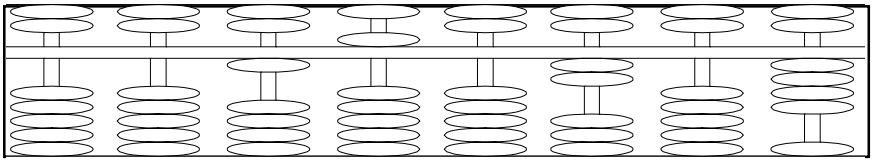

The Measure of Education:
A Review of the Tennessee Value Added Assessment System

A. Paige Baker
Senior Legislative
Research Analyst

Dengke Xu
Senior Legislative
Research Analyst

Ethel Detch, Director
Office of Education Accountability
1360 Andrew Jackson Building
500 Deaderick St.
Nashville, Tennessee 37243-0268
615/532-1111
Fax 615/532-9237

William R. Snodgrass
Comptroller of the Treasury

April 1995

Executive Summary

When the General Assembly passed the Education Improvement Act in 1992, legislators wanted a means by which the public could hold educators accountable for their performance. They adopted a model called the Tennessee Value-Added Assessment System, known as the TVAAS. Developed by Dr. William Sanders of the University of Tennessee, the model attempts to measure district, school, and teacher effect on student academic *gains*—rather than emphasizing achievement scores at a single point in time, as has been the generally accepted practice. In order to calculate academic gain scores, the TVAAS uses data generated by portions of a standardized test, the Tennessee Comprehensive Assessment Program (TCAP). (TCAP may be given in grades K-12, but TVAAS uses data generated from the test in grades 2-8.)

The Education Improvement Act requires the Commissioner of Education, with the approval of the State Board of Education, to place on probation systems and schools that are not making sufficient progress in: student value-added gains as calculated by the TVAAS; attendance rates; dropout rates; and other performance areas adopted by the State Board of Education. Although the TVAAS is just one of those areas, it has received the largest share of attention and discussion.

Through this study the Office of Education Accountability sought to evaluate the status of the model's implementation and to identify issues that state officials may need to address. Office of Education Accountability analysts and Dr. Sanders were unable to reach agreement on a number of issues surrounding the value-added assessment; his comments have been included in Appendix G. The study concludes:

Although the Office of Education Accountability has attempted to provide a general analysis of the TVAAS model, major components of the model need further evaluation. Such an evaluation might lay to rest many of the questions and concerns people have raised about the theoretical and statistical bases of the model. A number of statisticians and educational measurement experts interviewed for this report believe that the TVAAS is pioneer work in educational evaluation. Until qualified experts have the opportunity to examine all aspects of the model, however, the TVAAS will be hampered by questions about its validity. (See page 9.)

Sanders indicates that he welcomes such an evaluation, and he cites studies that have already been performed using information from the TVAAS. (See pp. 41-45.)

Because of unexplained variability in national norm gains across grade levels, it is not clear that those scores are the best benchmark by which to judge Tennessee educators. National norm gains vary greatly across grades and subjects—yet each district, school, and classroom is expected to achieve student academic gains equal to or greater than those national norm gains. The lower gain expectation for some grades appears to make it easier for those students to achieve 100 percent of national norm gains than students in other grades.

As an example, 7th graders are expected to gain only four points in social studies, while 8th graders are expected to gain 12 points. In fact, the 7th grade average gain score in social studies in Tennessee was *331.2 percent* of the national norm gain. These results

could be interpreted in different ways: perhaps the state is doing an excellent job of teaching 7th grade social studies in every system across the state—or perhaps the national norm score does not provide a realistic standard of comparison. (See pages 10-11.)

Sanders stresses that emphasis should by placed on district and school cumulative gains, rather than gains by grade by subject: "Extracting the scores for an individual subject in an individual grade distorts the meaning of the value-added assessment, since to achieve gains across grade levels requires a concerted effort of the entire faculty." (See pp. 45-46.)

There are large changes in value-added scores from year to year, and teachers and administrators have been unable to explain those variations. As a result, the model may not help identify superior educational methods to the extent policymakers had hoped. District and school cumulative value-added scores may be very different from one year to the next, without any obvious explanation. Some districts had cumulative value-added gain scores of 120 percent in one year and 60 percent in the next year. Some schools had 200 percent *cumulative* gains in one year and no gains—or even negative gains—in the next year. The variability in value-added scores raises concerns that will need to be addressed as implementation of the model continues. (See pages 11-14.)

Sanders states that the large variations in district and school value-added scores are to be expected in the early phases of the implementation of the model. He believes that much of the variation reflects either: the failure of schools and systems to adapt their teaching methods to children's instructional needs; or the positive responses of teachers and schools to the information supplied by the TVAAS. He believes the variations will decrease over time. (See pp. 46-50.)

The factors affecting student academic gain have not been identified, yet the model infers teacher, school, and district effect on student academic gain from the results of the value-added process. Education research has identified at least five factors that influence student learning including: individual student characteristics; student family characteristics; peer group variables; and teacher and school effects. The value-added model does not address student, family, or peer group variables, nor does it specify clearly how much of student gain is, or should be, attributable to a teacher in the classroom, or the school or district. (See pages 14-17.)

Sanders states that the statistical methodology in the value-added assessment model filters out the effects of many demographic characteristics on student academic gain. (See pp. 50-51.)

The case of the Scotts Hill School raises questions about using the TVAAS to evaluate school effect on student learning. Scotts Hill School is located on the county line between Henderson and Decatur counties, and it serves students from both counties. When Office of Education Accountability analysts compared Scotts Hill School of Henderson County with Scotts Hill School of Decatur County (which are really the same school) they expected the scores to be similar, if not the same. Instead, they discovered that the scores were quite different. Since the school and classroom conditions are the same for students

from both counties, Office of Education Accountability analysts question the divergence of the TVAAS scores. (See pages 17-18.)

Sanders states that the Office of Education Accountability's interpretation of this case is "simply wrong." He believes Scotts Hill School tailors its curriculum to lower-achieving students—more of whom come from Decatur County—and that this explains the differences in value-added scores between the two groups of students. (See pp. 52-55.)

The state needs to assure that VARAC's documentation practices are adequate and that there are plans in place to run the model in the event the present system is disabled. In addition, the state should assure that there is adequate dissemination of the data. The Office of Education Accountability is concerned that implementation of the TVAAS could be adversely affected if something were to happen to the computer system at VARAC or to key personnel. The possibility of sabotage or accidents underscores the need to disseminate the data more widely and to have other persons, possibly in the State Department of Education, who are trained to run the model. (See pages 18-19.)

Sanders states that written procedures for the value-added software exist, and that "operation of the value-added assessment program is not contingent upon the presence of any single person or group of persons." (See p. 55.)

The model is not easily explainable to those affected by it. The complexity of the model makes it difficult to convince those most affected by the evaluation that it is an appropriate and reliable means of measuring the impact educators have on student learning. Confidence could be fostered by a greater emphasis on positive ways that value-added assessment results could be used to enhance educational practices. (See page 19.)

Sanders agrees that the model is not easily explainable, and he cites the Value-Added Research and Assessment Center's efforts to provide information training on the use of value-added results to teachers and administrators. (See pp. 55-56.)

The "high stakes" nature of the TCAP test may create unintended incentives for both educators and students. Possible unintended incentives for teachers and administrators include the temptations: to "teach to the test"; to use instructional time to teach test-taking skills; or to engage in illegal test administration procedures. There may also be incentives to classify as many children as possible in special education categories (because special education value-added scores are not counted in TVAAS classroom evaluations), or to retain low achievers rather than promoting them to the next grade.

Students may also have an incentive to perform poorly in an attempt to damage a teacher's personnel evaluation. The ramifications of a testing system that places high stakes on teachers and no stakes on students needs to be studied. (See pages 19-21.)

Sanders believes it is unlikely that isolated examples of teacher or student misconduct would be sufficient to invalidate the results of the TVAAS. He also believes that "teaching to the test" is "extremely difficult if not impossible"—and that "there is no efficacy to [the] practice" of classifying low-ability or low-achieving students as special education students. (See pp. 56-57.)

Recommendations

The report recommends that all components of the TVAAS be evaluated by qualified experts knowledgeable of statistics, educational measurement, and testing.

The Department of Audit should perform an Information Systems Assessment to evaluate VARAC's documentation practices and assess the safety and security of the TVAAS. The state needs assurance that reasonable operational procedures are in place to protect the hardware, software, and data.

The State Board of Education and the State Department of Education need to identify unintended incentives for educators and students and consider ways to reduce their likelihood. (See pages 22-23.)

Table of Contents

Introduction

In 1992, members of the Tennessee General Assembly passed a major education reform initiative—the Education Improvement Act (EIA)—which mandated a number of significant changes for students, teachers, and schools. One of the EIA's most important sections describes the means by which the state plans to hold educators accountable for their performance. The primary component of the EIA's accountability section is the Tennessee Value-Added Assessment System (TVAAS).[1] This statistical model—designed to analyze the effects of districts, schools, and teachers on students' learning—was developed by Dr. William Sanders, a statistician with the University of Tennessee at Knoxville.

In the early 1980s, Sanders began to explore the feasibility of combining statistical techniques and measures of student achievement (i.e., test scores) to evaluate teacher and school influence on student learning. Using student test scores to evaluate teachers is not a new idea, but there have been many problems associated with using raw test data for this purpose. One major drawback has been that student academic achievement may be affected by factors over which schools and teachers have little or no control (e.g., socioeconomic status, parental education, and innate student ability).

One of the key claims for the value-added assessment model is that it overcomes the major drawbacks associated with using raw test data to evaluate districts, schools, and teachers. According to Sanders, the advantage of his model is that it focuses on student academic *gains* rather than raw achievement scores. Value-added assessment means that districts, schools, and teachers will be evaluated on the *improvement* students make, rather than on their absolute levels of achievement. It means that teachers and schools will be held accountable for taking students at whatever level they are achieving and helping those students progress each year.

Using gain scores also means that teachers, schools, and districts will not be penalized for the low raw achievement scores of students simply because those students started at disadvantaged positions. In this way, the model makes allowances for those educators who work with low-achieving students, while still demanding that they help those students achieve academically. The guiding principle of TVAAS is that "[s]ociety has a right to expect that schools will provide students with the opportunity for academic gain *regardless of the level at which the students enter the educational venue*. In other words, all students can and should learn commensurate with their abilities."[2]

From early tests of the model in the Knox County, Blount County, and Chattanooga City school districts, Sanders concluded that:

1. There were measurable differences among schools and teachers with regard to their effect on indicators of student learning.

---

[1] According to the statute, TVAAS is "a statistical system for educational outcome assessment which uses measures of student learning to enable the estimation of teacher, school, and school district statistical distributions." *TCA* §49-1-603 (a) (1).

[2] William L. Sanders & Sandra Horn. *The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment*, 1994. (Emphasis added.)

2. The estimates of school and teacher effects tended to be consistent from year to year.
3. Teacher effects were not site specific, i.e., a gain score could not be predicted by simply knowing the location of the school.
4. Student gains were not related to the ability or achievement levels of the students when they entered the classroom.
5. The estimate of school effects was not related to racial composition of the student body.
6. There was very strong correlation between teacher effects as determined by the data and subjective evaluations by principals and supervisors.[3]

He was convinced that these findings indicated that *gains*, rather than raw achievement scores, were a legitimate way of determining the influence of districts, schools, and teachers on student learning.[4] Sanders uses a "mixed model" statistical methodology to estimate that influence.

Methodology

In accordance with its mission as determined by *TCA* §4-3-308, the Office of Education Accountability elected to perform an assessment of the Tennessee Value-Added Assessment System, and to evaluate the implementation of the model. The goal of the evaluation was to answer the following questions:

- What is the value-added assessment model, and how does it work?
- How useful is the model and what are its limitations?
- Is implementation proceeding in the way the General Assembly directed?

In the course of the evaluation, analysts performed a review of the TVAAS data for both districts and schools; surveyed literature on both statistical modeling and educational assessment; and interviewed experts in both fields.[5] They also interviewed Dr. Sanders and members of his staff; personnel from the State Board of Education[6]; personnel in the State Department of Education; and representatives from the Tennessee Education Association (TEA), the Tennessee School Board Association (TSBA), and the Tennessee Organization of School Superintendents (TOSS).[7] A spirited debate on the Internet's Education Policy Analysis bulletin board (in which Sanders and his staff participated) was also helpful.

---

[3]William L. Sanders and Sandra Horn. *An Overview of the: Tennessee Value-Added Assessment System (TVAAS) with Answers to Frequently Asked Questions*. Knoxville: University of Tennessee, p. 2.

[4]In this context, the term "district" refers to all 139 school systems.

[5]Some of those interviewed were suggested by Dr. Sanders. Others were suggested by various sources within Tennessee state government and members of the academic community.

[6]Karen Weeks, Research Analyst for the State Board of Education, deserves a special note of thanks. She provided a great deal of information and very helpful suggestions on a draft of this report. A number of her suggestions have been included in the final product.

[7]For a complete list of those interviewed, see Appendix A.

Educational Assessment in Tennessee

The Tennessee Comprehensive Assessment Program (TCAP) includes a series of standardized tests that serve as the foundation for the accountability portion of the EIA. The TCAP tests have been administered annually by the Tennessee Department of Education since 1990.[8] The TCAP includes four types of tests:

1) A norm-referenced test. This type of test makes it possible to compare the performance of Tennessee students to student performance nationwide. The test is given to students in grades 2-8 and grade 10. Results of this portion of the test in grades 2-8 are used in the value-added assessment.

The State Department of Education contracts with the CTB/McGraw-Hill Company for the Comprehensive Tests of Basic Skills, Fourth Edition (CTBS/4). CTBS/4 measures students' academic skills in common subject areas of reading, language arts, mathematics, science, and social studies.[9] Test developers review curriculum guides and text series nationwide to identify educational objectives within broad categories.

2) A criterion-referenced test. This test is constructed specifically for Tennessee students based on the Tennessee public school curriculum. It is administered with the norm-referenced test in grades 2-8 to measure student mastery of the math and language arts curricula. The results of this portion of the TCAP are *not* used in the value-added assessment.

The combination of the norm-referenced test and the criterion-referenced test is referred to as the *Tennessee Achievement Test*.

3) The TCAP Proficiency Test. This test measures the achievement of basic math and language arts skills by high school students.[10] The results of this portion of the TCAP are *not* used in the value-added assessment.

4) A writing assessment. In fall 1994, the writing assessment was administered statewide for the first time in grades 4, 8 and 11. The results of this portion of the TCAP are *not* used in the value-added assessment.

---

[8]The numbers of students tested were as follows:

1990—450,638
1991—492,387
1992—496,897
1993—504,478
1994—508,228

[9]In the spring of 1988, McGraw-Hill tested a sample of 156,000 students in grades K-12; their scores were then used to standardize the test and to establish a national norm. Those students were drawn from 148 public school districts, 18 Catholic dioceses, and 62 private, non-Catholic schools stratified by region, community type, and size.

[10]This test has recently been replaced by the new Tennessee Student Competency Test. The Competency Test will be given to all incoming freshmen this year. The Proficiency Test will be given to current seniors, juniors, and sophomores; the latter two groups will continue to take the Proficiency Test until they graduate.

Steps in the Value-Added Evaluation

The current TVAAS analyzes academic gains in five subject areas: math, reading, language arts, social studies, and science for grades 3-8.[11] Students take the TCAP tests during a specified period of time in late March or early April; the scoring sheets are then sent to the State Testing and Evaluation Center (STEC) in Knoxville.[12] When the tests have been scored and the data collected, the information is given to Sanders and his staff at the University of Tennessee Value-Added Research and Assessment Center (VARAC).

VARAC has created a database that merges new student test data with student data from previous years. This merging process enables VARAC to track student academic progress over time. The database contains the student file, the teacher file, and the school name file; the files are linked to each other by student identification numbers or district/school codes.[13]

Merged data provide more information than a number of single year data could provide because the process of merging connects individual years of data. These connections provide information about student academic progress that would not be easily uncovered if one looked merely at individual years of data. The TVAAS model uses all the information available to estimate gains for districts, schools, and teachers, rather than for individual students.

A database containing such a large number of records presents certain problems, however. There are discrepancies in the data or missing data for a variety of reasons: students repeat or skip grades; students give middle initials on the coding sheet one year, but not the next; students change districts or classrooms. As a result, there are incomplete records. According to John Schneider and Paul Wright, data analysts for the project, the student data match-rate is about 90 percent.[14]

To improve the quality of matching, each year VARAC analysts edit the new data prior to matching. A weighting process is used to compensate for missing information; complete student records are weighted more heavily than incomplete records. Weighting factors for incomplete student records are dependent on the amount of information provided by available records.

---

[11]The Education Improvement Act requires that all tests used in TVAAS "shall be fresh, nonredundant equivalent tests, replaced each year." *TCA* §49-1-610. The contract between CTB/McGraw-Hill and the State Testing and Evaluation Center (Contract #DG940529) specifies the criteria for the selection of the norm-referenced test items in TCAP: 'from the CTBS/4 Survey A and B (25-33%) plus selections from CRT items from CTBS/4 and CAT/5 Complete Batteries (75%).' The actual tests provided by CTB/McGraw-Hill contain approximately 70 percent new items and 30 percent repeated items—used somewhere in the past, not necessarily the previous year in Tennessee.

[12]The testing dates: 1990—April 9-27; 1991—April 8-19; 1992—April 6-17; 1993—March 29-April 16; 1994—March 28-April 15; 1995—March 27-April 14. One of the major complaints raised by teachers about the value-added assessment is that their students are being tested too early in the year. They assert that the TCAP tests are covering material they have not had time to teach.

[13] The program to handle this matching was written on SAS® software. An IBM RISC System/6000 model 580 computer with one gigabyte of RAM is dedicated for the operation.

[14]A 90 percent match rate is considered excellent—even though the practical effect is that there are approximately 40,000-50,000 student data files that cannot be matched each year.

Some Illustrations

The value-added scores in the 1993 and 1994 *Report Cards*[15] were cumulative gains of grades 3-8, expressed as percentages of the national norm gains. The value-added scores represent the average of the most current three years.[16]

- A district's value-added score equals:
  District's grades 3-8 three year average cumulative point gain %
  national norm point gain.

- A school's gain score equals:
  School's grades 3-8 three year average cumulative point gain %
  national norm point gain.[17]

Some examples of actual district and school "report cards" appear in Appendix G.

Implementation of Accountability Measures

The accountability provisions of the EIA require:

- Value-added assessment for grades 3-8 at the district, school, and classroom levels;
- Value-added assessment for high school students in academic subject areas[18];
- Performance standards;
- Fiscal accountability standards for districts;
- Annual report on standards;
- Incentive funding for schools;
- Probation for school districts;
- Guidelines for use of student performance data in educator evaluation.[19]

In September 1993, the State Department of Education released the first *Report Card* containing value-added assessments for grades 3-8 at the district level. The following year, the Department released the first school-level evaluations. Classroom evaluations are scheduled to be produced in the summer of 1995.

The Commissioner of Education has recommended and the State Board of Education has approved performance goals in several areas: academic gain, attendance, dropout rate, and promotion.[20]

---

[15]These reports are issued jointly by the Governor's Office and the Tennessee Department of Education.

[16]See the appendix for value-added scores in the 1993 and 1994 Report Cards for school systems.

[17]Some schools have only some of the grades 3-8. Their evaluations are adjusted accordingly.

[18]Tennessee has recently signed a contract with CTB/McGraw-Hill to develop the subject matter tests for high school students. The mathematics tests—the first to be developed—are scheduled to begin in spring 1996. Other academic subjects are to be phased in over a period of time; the phase-in is to be completed by spring 1999.

[19]This information was taken from a handout prepared by the staff of the State Board of Education for the Senate Education Committee meeting of February 1, 1995.

[20]In July 1992 the Commissioner of Education recommended and the State Board of Education approved *five* performance goals. The fifth goal, proficiency rate, will not be used for three years because the proficiency test has been upgraded to a new competency test and at least three years of data are needed in order to measure progress towards the goal.

- **Academic Gain:** As established by the Education Improvement Act, average gain must be greater than or equal to the national norm gain in each of the five subject areas (math, reading, language arts, social studies, and science).
- **Attendance:** Attendance of K-6 must be greater than or equal to 95 percent. Attendance in grades 7-12 must be greater than or equal to 93 percent.
- **Promotion:** For grades K-8, the goal of promotion is 97 percent.
- **Dropout Rate:** The goal for dropout rates in grades 9-12 is 10 percent or less.

To ensure that districts would meet the goals by the year 2000, the State Department of Education recommended and the State Board of Education approved annual standards. Each standard is determined by taking the difference between the district's current performance and its goal and dividing the difference by the number of years remaining until the year 2000. The *State of Tennessee Report Card* details the progress each district is making toward meeting its goals.

As noted above, classroom value-added results will be calculated in the summer of 1995. By law, classroom value-added scores are protected information, and will not be released to the public. They will be released "only to the specific teacher, the teacher's appropriate administrators as designated by the local board of education, and school board members."[21] The *TCA* states:

> The [State] board shall develop, and provide to local education agencies, guidelines and criteria for the evaluation of all certificated persons employed by such agency; provided, however, that such mandatory criteria shall include, but not be limited to:
>
> (1) Classroom or position observation followed by written assessment;
> (2) Review of peer evaluations;
> (3) Personal conferences to include discussion of strengths, weaknesses, and remediation; and
> (4) Other appropriate criteria including the Sanders model, related to the responsibilities of the employee.[22]

The Department of Education is in the process of revising the State Model for Local Evaluation. This model, which provides guidelines for local evaluation of educators, will be revised to include guidelines for use of classroom value-added results in the local evaluation of teachers.

The *TCA* states that TVAAS data may not be used in teacher evaluations until three years of data are available; the third year of value-added data will be available the summer of 1995.[23] Student scores will be counted in teacher evaluations only if they have been

[21] *TCA* §49-1-606(b).

[22] *TCA* §49-1-302(d).

[23] Three years of data will produce only two gain scores, however. Representatives from both the State Department of Education and the State Board of Education indicated that gain scores ought not to be used for the purpose of personnel evaluation until another year of value-added data could be collected.

present in class for a specified number of days, and special education students are excluded from the classroom value-added assessment.[24]

Consequences for Districts and Schools

Value-added assessment is used in combination with other performance indicators (attendance rates, dropout rate, and promotion rate) in determining both rewards for schools and sanctions for districts. Individual schools are eligible for rewards if they meet all performance goals that apply to the grade configuration of their school; districts are subject to sanctions if they fail to meet performance standards regarding academic gain and other performance indicators.

Individual schools must have a cumulative value-added score equal to or greater than 100 percent of the national norm gain in each of the five subject areas. Each school that meets or exceeds the specified goals is rewarded with a share of available incentive funds. The first incentive funds were awarded in fall 1994; a total of $500,000 was awarded to 120 schools that met or exceeded their goals.

Districts must achieve at least a 95 percent cumulative gain average over all five subject areas, with a minimum cumulative gain of 90 percent in each subject. If districts fail to achieve this level of gain, they must demonstrate that they are making statistically significant progress toward reaching that level. Those that cannot demonstrate acceptable progress are subject to sanctions.

Beginning in October 1995, the Commissioner of Education, with the approval of the State Board of Education, may place any district or school failing to meet the required performance standards on probation. During the first year of probation, the State Department of Education will conduct a study and make recommendations for improvement. If a district or school remains on probation for two consecutive years, the commissioner may recommend to the Board that the superintendent and members of the local board of education be removed from office.[25]

Cost of the Testing and Value-added Assessment Programs

According to figures provided by the State Department of Education, the budget for the 1994-95 TCAP testing program in grades 3-8 is $2 million. In fiscal year 1994, the state allocated $250,000 for the development of high school subject tests. The fiscal year 1995 budget includes $2.7 million in new dollars for the development of the high school subject tests in five areas.

---

[24]Some special education students take the TCAP, however, and their scores will be counted in the school and district—though not the classroom—reports.

The *TCA* states: "A specific teacher's effect on the educational progress of students may not be used as a part of formal personnel evaluation until data from three (3) complete academic years are obtained. Teacher effect data shall not be retained for use in evaluations for more than the most recent five (5) years. A student must have been present for one hundred fifty (150) days of classroom instruction per year or seventy-five (75) days of classroom instruction per semester before that student's record is attributable to a specific teacher. Records from any student who is eligible for special education services under federal law will not be used as part of the value-added assessment." *TCA* §49-1-606(a).

[25]*TCA* §49-1-602.

In addition, the Value-Added Research and Assessment Center received the following amounts:

1992-93—$314,500
1993-94—$353,000
1994-95—$373,400

The State Department of Education projects that VARAC will receive the following amounts in the next three years:

Projected 1995-96—$370,400
Projected 1996-97—$370,000
Projected 1997-98—$370,000

Conclusions

- **Although the Office of Education Accountability has attempted to provide a general analysis of the TVAAS model, major components of the model need further evaluation.** ***(See response on pp. 41-45.)***

Additional evaluation of the value-added assessment model might lay to rest many of the questions and concerns people have raised about the TVAAS—but, as yet, no such comprehensive evaluation has been performed. Sanders indicates that he welcomes such an evaluation, although he has been extremely protective of both the data and computer software used to run the value-added calculations. While Sanders has made the *Report Card* results available in different forms, he apparently has not provided complete information to anyone who could replicate the model.[26] He indicates that he is concerned about contractual obligations and copyright infringement, but it would be difficult, if not impossible, to perform an adequate evaluation without access to the student database and the software.

Access to the database and the value-added software is guaranteed by the state's contract with the Value-Added Research and Assessment Center. The state contract with VARAC says:

> The software developed under this project will be copyrighted by the University of Tennessee, subject to the following conditions. The State shall be granted an unrestricted license to use for its own purposes, modify, duplicate, and/or distribute to any Tennessee public school system the developed software, entirely at the State's discretion and without any further payment, fees, or royalty beyond the terms of this contract.[27]

The State Department of Education needs access to the student database and the value-added software in order to replicate VARAC's results. For purposes of evaluation, reputable experts cannot substantiate VARAC's claims without access to a great deal more information than is currently available.

The recent debate on the Internet's Education Policy Analysis Archives indicates that the basic statistical procedures involved in the model are understood by many researchers. Many experts in the fields of statistics and educational measurement have expressed an interest in the model, and a desire to analyze the process and the results of value-added assessment—but they have been frustrated by the lack of available information. A number of those interviewed for this report believe that the TVAAS is pioneer work in educational evaluation—but they would like to examine both the statistical work and the assumptions and policy implications of the model.

The issues associated with the value-added assessment model are not merely statistical ones, but also include issues of educational measurement and public policy. Both statisticians and educational measurement experts need the opportunity to test the model. Without further evaluation, the state—and its educational practitioners—cannot determine the validity of the value-added model.

---

[26] Replication of experiments is accepted practice in most academic disciplines; replication is used to strengthen arguments for or against the validity of a particular model or theory.

[27] Contract #ID-5-05479-5-00.

- **Because of unexplained variability in national norm gains across grade levels, it is not clear that national norm gains are the best benchmark by which to judge Tennessee educators. *(See response on pp. 45-46.)***

National norm gains vary greatly across grades and subjects—yet each district, school, and classroom is expected to achieve student academic gains equal to or greater than those national norm gains. It may be reasonable to assume that students in the lower grades will, in general, gain more than their counterparts in the upper grades. What is difficult to justify are the differences in expectations among adjacent grade levels. The chart below demonstrates the differences between the national norm gains and Tennessee gains:

1994 TVAAS State Summary Report							
Three-year-average gain by grade and subject							
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Cumulative. Gain
Math							
National Norm Gain	60.0	26.0	25.0	19.0	15.0	18.0	163.0
TN Math Gain	48.4	28.3	23.4	20.5	16.8	18.8	156.2
TN% of National Norm Gain	80.6%	108.8%	93.6%	108.1%	112.3%	104.6%	95.9%
Reading							
National Norm Gain	30.0	21.0	21.0	18.0	9.0	10.0	109.0
TN Reading Gain	27.9	23.0	22.9	14.9	14.5	10.9	114.1
TN% of National Norm Gain	93.0%	109.3%	109.1%	82.6%	161.1%	109.2%	104.7%
Language							
National Norm Gain	29.0	11.0	17.0	15.0	10.0	8.0	90.0
TN Language Gain	18.8	15.2	22.9	11.1	11.9	16.6	96.5
TN% of National Norm Gain	64.7%	138.5%	134.6%	74.2%	119.2%	207.8%	107.3%
Social Studies							
National Norm Gain	39.0	22.0	22.0	10.0	4.0	12.0	109.0
TN Social Studies Gain	33.2	21.6	17.0	5.4	13.2	11.7	102.1
TN% of National Norm Gain	85.2%	98.0%	77.1%	53.6%	331.2%	97.6%	93.7%
Science							
National Norm Gain	35.0	19.0	23.0	13.0	11.0	9.0	110.0
TN Science Gain	26.2	28.2	14.4	11.6	14.4	12.9	107.7
TN% of National Norm Gain	74.8%	148.5%	62.7%	89.4%	131.2%	143.5%	98.0%

For instance, on the language portion of the norm-referenced test:

- 3rd graders are expected to gain 29 points;
- 4th graders are expected to gain 11 points; and
- 5th graders are expected to gain 17 points.

There is no readily apparent explanation for this wide range of norm gain scores—yet teachers, schools, and districts will be required to meet these gain levels in order to avoid sanctions.

Another example of variability is evident on the social studies portion of the norm-referenced test. Sixth graders are expected to gain 10 points, and 8th graders are expected to gain 12 points—yet 7th graders are expected to gain only four points. In Tennessee, the last three-year-average gains in social studies were:

- 5.4 points for 6th graders;
- 13.2 points for 7th graders; and
- 11.7 points for 8th graders.

This means that in social studies:

- 6th graders achieved only 53.6 percent of the national norm gain;
- 7th graders achieved an astounding 331.2 percent of the national norm gain; and
- 8th graders achieved only 97.6 percent of the national norm gain.

Does this mean Tennessee schools and teachers of 6th graders did a very poor job, and schools and teachers of 7th graders did an extremely good job in terms of student learning? Not necessarily. Because of low expectations—based on the national norm gains—for 7th graders, Tennessee 7th graders, on average, achieved more than 100 percent of the national norm gains *in all five subject areas*. Across all subjects, the *lowest* 7th grade statewide average gain score was 112.3 percent of the national norm gain in math. The low expectations for 7th graders simply make it much easier to achieve the national norm gain level.

Conversely—because of high expectations for 3rd graders—Tennessee 3rd graders, on average, had less than 100 percent of national norm gains in all five subject areas. The *highest* 3rd grade statewide average gain score was 93 percent of the national norm gain in reading.

The statewide three-year-average gain figures indicate that it would be much easier to achieve the national norm gain in the 7th and 8th grades than it would be to achieve the national norm gain in any of the elementary grades. Thus, using national norm gains as the benchmark for Tennessee educators puts those schools serving only elementary students in a disadvantaged position.

- **There are large changes in value-added scores from year to year, and teachers and administrators have been unable to explain those variations. As a result, the model may not help identify superior educational methods to the extent policymakers had hoped.** ***(See response on pp. 46-50.)***

District and school cumulative value-added scores may be very different from one year to the next, without any obvious explanation for the differences. This makes it difficult for

educators and policymakers to determine what value-added scores really mean—or how educational practice needs to be changed to increase student academic gains.

The district-level data given in the *Report Cards* are three year *average* data. This means that only one year of the three years of data is replaced annually—so only one third of the information is new from one year's *Report Card* to the next. Therefore, changes in cumulative gain scores from year to year are actually *much larger than they appear to be in the Report Card*. Using the three year average cumulative gain score helps to even out high and low gain scores—but the three year average also obscures the wide variations in gains from year to year. The three-year average is not particularly helpful for those who are looking for explanations for changes in cumulative gain scores.

Individual year data at the district and school levels is much less consistent than the three-year average given in the *Report Cards*. The Memphis City School District is a good example of a district with significant changes in its cumulative gain scores from year to year. In the 1994 *Report Card*, the district had three-year average gain scores of 84.0 percent in math, 90.4 percent in reading, and 84.2 percent in language arts. The three individual year scores show large differences in scores from year to year.[28] The variation in gains reaches as much as 40.2 percentage points (in Language) from one year to another.

Year	Math	Reading	Language
1991[29]	100.1	105.4	118.3
1992	93.9	94.0	96.9
1993	91.6	109.9	98.0
1994	66.5	67.5	57.8
3 yr. avg. 1993[30]	95.2	103.1	104.4
3 yr. avg. 1994[31]	84.0	90.4	84.2

There are a dozen districts that had average double-digit changes (positive or negative) in their three-year average cumulative gain scores from the 1993 *Report Card* to the 1994 *Report Card*.[32] It is important to remember that these are *percentage* changes, and that they are cumulative for an entire district's 3rd-8th grade populations.[33] One would expect

[28]Source of information: Memphis City *Report Card Supplement*, October 1994

[29]It should be noted that the 1990 TCAP—upon which the 1991 gain scores were based— was given to only a *sample* of students in the Memphis City School District. Sanders believes that this made gain scores in 1991 artificially high for that district.

[30]This information was taken from the 1993 *Report Card*.

[31]This information was taken from the 1994 *Report Card*.

[32]Seven districts had double-digit negative average changes: Memphis City (-13.2%); Greeneville City (-12.03%); Morgan County (-11.48%); Covington City (-11.00%); Dayton City (-10.94%); Tipton County (-10.16%); and Meigs County (-10.06%). Five districts had double-digit positive average changes: Bells City (20.04%); McKenzie Special School District (13.51%); McNairy County (11.47%); Etowah City (11.02%); and Hardin County (11.02%).

[33]There are a few districts that consist of a single elementary school. In this case, scores are likely to be less stable. See discussion following.

that these wide variations in gains should be explained by some significant changes at the district level, but evidence of such changes was not apparent. Even major changes at the district level—such as the arrival of a new superintendent—would not be expected to generate that kind of impact in the short term.

At the school level, the cumulative gain scores are understandably less stable than district gain scores because the number of students in a single school is much smaller; the smaller the number of students in the sample, the more volatile the gain scores will be. There are schools, however, that had 200 percent cumulative gains in one year and no gains—or even negative gains—in the next year. Oakdale School in Morgan County provides an example of wide swings from year to year:

1994 TVAAS Report

School: Oakdale School (Morgan County)% Norm Gain (Cumulative Grades 3 to 8)

Year	Math	Reading	Language Arts	Social Studies	Science
1992	125.2	128.2	107.7	100.6	108.4
1993	64.6	100.9	103.3	59.6	82.6
1994	149.4	65.4	55.3	101.6	108.6
3 Yr Avg.	113.1	98.2	88.8	87.3	99.9

- For all grades 3-8, cumulative gain scores in math dropped 60.6 percent from 1992 to 1993 and rebounded 84.8 percent in 1994;
- For all grades 3-8, cumulative gain scores in reading dropped 27.3 percent from 1992 to 1993 and dropped again in 1994 by 35.5 percent;
- For all grades 3-8, cumulative gain scores in language arts fell 48 percent in one year;
- For all grades 3-8, cumulative gain scores in social studies dropped 41 percent from 1992 to 1993 and rebounded 42 percent in 1994;
- For all grades 3-8, cumulative gain scores in science dropped 25.8 percent from 1992 to 1993 and rebounded by 26 percent in 1994.

The Bells Elementary School provides a similar example of a school with large variations in its gain scores:

1994 TVAAS Report

School: Bells Elementary (Crockett County)% Norm Gain (Cumulative Grades 3-5)

Year	Math	Reading	Language Arts	Social Studies	Science
1992	75.5	129.5	122.8	73.9	55.9
1993	68.7	62.7	80.0	71.9	56.8
1994	105.8	131.3	122.0	105.8	154.7
3 Yr Avg.	81.6	117.4	119.2	91.3	94.4

- For grades 3-5, cumulative gain scores in math increased 37.1 percent in one year;
- For grades 3-5, cumulative gain scores in reading dropped 66.8 percent from 1992 to 1993 and rebounded by 68.6 percent in 1994;

- For grades 3-5, cumulative gain scores in language arts dropped 42.8 percent from 1992 to 1993 and rebounded by 42 percent in 1994;
- For grades 3-5, cumulative gain scores in social studies increased 33.9 percent in one year; and
- For grades 3-5, cumulative gain scores in science increased 97.9 percent in one year.

There are numerous examples across the state of schools that have experienced these kinds of dramatic changes from year to year. The less consistent the value-added scores, the less reliable the results will be—and the less confidence one can have in the conclusions one draws from those results. *Reliability* is a statistical term that describes the degree to which test scores are free of errors of measurement—that is, the degree to which they are consistent, dependable, or repeatable. One way to determine this is to observe reliability coefficients for the tests.[34] *The choice of an acceptable reliability level is a policy decision, however, not a statistical one.*

Reliability is an important concept, but *validity* is the most important consideration in test evaluation. Validity is the extent to which a test or measurement system measures that which one wants measured. In this case, there are two different questions about validity:

- Does the TCAP test measure what educators, policymakers, and the public want it to measure?
- Are the results produced by the value-added assessment model meaningful?

Statements about validity should refer to the validity of particular inferences from a test, not the test itself—because a test or measurement system that is valid for one purpose may not be valid for another. Test validation is the process of accumulating evidence to support such inferences.

If educators are unable to identify the factors causing wide variation in value-added scores, they will be unable to determine just how to improve instruction. The variability in scores raises concerns that will need to be addressed as implementation of the model continues.

- **The factors affecting student academic gain have not been identified, yet the model infers teacher, school, and district effect on student academic gain from the results of the value-added process. The validity of that inference needs to be evaluated.** ***(See response on pp. 50-52.)***

It may be reasonable to assume that the largest percentage of student gain in a given year is attributable to teachers, but classroom-level student gain may not be *equivalent* to a teacher's effect on students in the classroom. Student learning is a result of a complicated process. There are many factors, both in and out of classrooms that combine to produce the actual learning of students—teachers are only one factor contributing to student academic progress or lack thereof.

[34]Benchmark test-retest reliability coefficients for CTBS/4 vary from a low of 0.55 for Grades 11-12 science to a high of 0.92 for Grade 2 reading. A reliability coefficient of 0.55 is usually considered to be poor, while a reliability coefficient of 0.92 is generally thought to be excellent.

A large body of education research has identified at least five factors that influence student learning. Those factors include individual student characteristics (sex, age, preschool experience, affective and cognitive variables); student family characteristics (family size, parental education, parental occupation, and income); peer group variables (social class composition, racial composition, ability composition); teacher effect (teacher education, teacher experience, teacher ability, and personal characteristics); and school effect (expenditure per pupil, size, physical conditions, learning resources).[35] A study of the nation's eighth graders in nearly 1,000 schools in 1988 found that:

- There is a relationship between tested achievement and time spent doing homework and reading or watching TV.
- Parental involvement in a child's education is a crucial factor in promoting academic achievement.
- Eighth grade students who are 16 or older (students who have likely been held back and are over-age for grade) consistently score lower [on standardized achievement tests] than those who are 15, 14, or younger.
- There are relatively large mean differences on all tests in favor of white and Asian students compared to black and Hispanic minorities.
- Large group differences are seen between the groups of students from families in the top and bottom quartiles of socioeconomic status.[36]

The value-added assessment model does not include any of these student or family factors—nor does it specify clearly how much of student gain is, or should be, attributable to a teacher in the classroom. The model seems to assume that all gain (or lack thereof) is purely teacher-related, while it has not provided adequate evidence to support this contention.

Sanders believes that the value-added assessment model blocks the effects of many factors that appear to affect student achievement—including race, gender, socioeconomic status, and a number of others—by using each student as his or her own "control." A student's race and gender certainly remain the same from year to year, and Sanders believes that most family characteristics (income, level of parents' education, etc.) remain the same from year to year as well. By using longitudinal data on individual students to calculate value-added scores, Sanders believes that the effect of these non-educational variables on the value-added results are held constant, and thus can be minimized—and possibly eliminated entirely. "By [statistically] filtering out [the extraneous influences], you are left with influences that teachers have control of," Sanders explains. "We will partition those environmental controls from the teacher effects."[37]

Sanders says that the district level cumulative gains for each of the five subjects are uncorrelated with the percentage of students receiving free and reduced-price meals within the districts; neither are they correlated with the racial composition of the student body

---

[35] R. Gary Bridge, et al. *The Determinants of Educational Outcomes*. Cambridge, Massachusetts, 1979.

[36] Donald Rock and others. *The Tested Achievement of the National Education Longitudinal Study of 1988 Eighth Grade Class*. U.S. Department of Education, Office of Educational Research and Improvement, Education Information Center, 1991.

[37] *TEA NEWS*, September 1994, p.11.

within districts. This would support his earlier findings that cumulative gain scores were not related to student family socioeconomic status or racial composition of the student body. However, the conclusion was based on a simple linear test. There appears to be no *linear* relationship between cumulative gains and student family background or racial composition of the student body—but this does not completely rule out a relationship.[38] Further testing needs to be done.

More evidence is also needed to support Sanders' earlier finding that student gains are not related to student abilities. This finding appears to contradict the philosophical underpinning of the TVAAS: "all students can and should learn *commensurate with their abilities*."[39] No one would dispute that students do not possess equal learning ability. If all students learn commensurate with their abilities, they not only achieve at different levels, but may also learn at different rates. Even if all students were given equal educational opportunities, they might learn differently, both in terms of absolute achievement levels and academic gains.

This issue may not be a big concern in assessing schools or districts, because it is unlikely that all students with higher abilities would be in one school or one district and all students with lower abilities in another. Because of the small number of students in a classroom, however, it is a concern at the classroom (teacher) level. There could be a significant difference in student learning abilities from one classroom of students to another.

It is important to distinguish between students with lower learning *abilities* and those whose *achievement levels* are low—even though the two are highly correlated. Low achieving students may just need the "right" kind of teaching to spur them to higher levels of academic gain —and the TVAAS is designed to provide some indication of how well teachers are reaching those students. Students with lower learning abilities, however, may not be able to achieve the same kinds of gains that other students do. Every student *can* learn and deserves an equal opportunity to learn, but it may be unreasonable to expect every student to learn at the same rate—and to expect teachers to have an equal impact on every student's academic gain. Without taking account of the differences in student ability, estimates of teacher influence on student gains in learning could be biased against teachers having larger numbers of students with lower learning ability.[40]

The issue of students with lower learning abilities arises in conjunction with the exclusion of special education students' gain scores from the component of TVAAS that will evaluate teachers. According to the EIA, "records from any student who is eligible for special education services under federal law will not be used as part of the value-added assessment" of teachers.[41] This exclusion implies that value-added scores are partially

---

[38] A linear relationship is one in which the dependent variable changes in direct proportion to changes in the independent variable(s). An example would be the relationship for an hourly worker between the number of hours worked and the total amount earned. This is an example of a perfect linear relationship—there are many relationships that are linear, but not perfectly so.

[39] *The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment*, Sanders & Horn, 1994, emphasis added.

[40] This issue was suggested by the Internet debate on the Education Policy Analysis Archives.

[41] *TCA* §49-1-606(a).

influenced by the kind of students teachers have—otherwise this proviso would be unnecessary.

Sanders says that the TVAAS model is capable of accommodating the inclusion of variables that would filter out influences outside the control of educators, if such variables were found to be relevant:

> [I]t may be found that some socioeconomic confoundings could surface in the future which would necessitate the inclusion of appropriate covariables in the mixed model equations. Current findings suggest that the number of any needed covariables will be relatively small, if any; however, TVAAS readily accommodates such inclusion.[42]

If there *are* significant covariables—and the value-added model assumes none—the practical result is that the model could evaluate some teachers more favorably than others—depending on the student population in a given classroom, resources available to the teacher, etc. The question one must ask is: should the value-added assessment model assume there are no covariables because current analyses, based on a limited amount of data, have yet to uncover them? Or should VARAC assume there *are* covariables and actively search for evidence to the contrary?

- **The case of Scotts Hill School raises questions about the validity of using the TVAAS to evaluate school effect on student learning.** ***(See response on pp. 52-55.)***

Office of Education Accountability analysts discovered an interesting case in a school in the western part of the state. Scotts Hill is located on the county line of Henderson and Decatur counties. On paper, there are two schools: one is Henderson County's Scotts Hill School and the other is Decatur County's Scotts Hill School. Practically speaking, however, there is only one school, and students from both counties are treated identically. The principal divides students into four groups by gender and last year's TCAP scores; then each classroom is assigned relatively equal numbers of students from all four groups (males, females, above-average TCAP scores, and below-average TCAP scores).

The TVAAS calculates value-added scores for the "two" Scotts Hill Schools separately. The model calculates one set of gains for the Scotts Hill School of Henderson County, based only on the students from Henderson County. The model calculates another set of gains for Scotts Hill School of Decatur County, based solely on the students from Decatur County. Since the two schools are separate only on paper, one would expect that the estimates of school influence on student learning for the two groups of students would be similar.

The gains estimated for the school based on the two groups of students were quite different, however. Here is a set of actual data from the 1994 *Report Card* supplement for the two schools:

Estimated Three-year-average Gains and Standard Errors in Language

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	
Scotts Hill	24.5Y	2.1R	39.3G	5.4R	14.7G	18.3G	**Gains**
Decatur	4.8	5.1	5.3	4.9	5.0	4.6	**Std. Error**

[42] Dr. William L. Sanders. *TVAAS Seminar for the Tennessee State Department of Education*. February 17-18, 1994. Quoted from draft paper, p. 19.

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	
Scotts Hill	16.8R*	15.7G	26.5G	-1.9R*	7.6Y	27.7G	Gains
Henderson	4.5	4.5	4.3	4.2	4.6	4.2	Std. Error

Gain Diff.	7.7	13.6	12.8	7.3	7.1	9.4
t-score	1.6	2.7	2.4	1.5	1.4	2.0

The *t*-score in the preceding table is simply the difference of gains between the two schools divided by the larger standard error of the two reported (which is from Scotts Hill of Decatur).[43] The general rule is: if the *t*-score is larger than two, the difference is statistically significant—which means the probability that the difference is due to chance is very small. Three of the six differences are statistically significant.[44]

The important issue is how one explains the significant differences. In reality, there is one school with two sets of estimates of its effects on student learning. According to school officials, both the students from Henderson County and the students from Decatur County are treated impartially with respect to classroom assignments. Students from the "two schools" actually attend the same school and have the same teachers. Therefore, the school and classroom effects on those students should be similar—*unless* there is some factor affecting student achievement that has little or nothing to do with the school or the teachers.

One possible explanation is that the two groups of students are not equal and they contribute to the gains differently— that school level student academic gains are affected not only by *school* influence on student learning but also by other influences not specified in the value-added model. If that is the case, then indicators of school effects on student learning are dependent on factors other than teachers and schools—such as students themselves—and the "blocking" effect in the model is insufficient.

Sanders believes that schools and teachers cause gains (or lack thereof); in this case, he believes the differences in gain scores between the two groups of students can be attributed to Scotts Hill's targeting teaching strategies at lower-achievers. Office of Education Accountability analysts believe this case may suggest otherwise—that the differences in test scores between these "two" schools are caused by differences in students themselves. In a sense, this case "controls for" the effects of schools on student performance. The differences in student gain scores cannot be attributed to the school because the school is the same for both groups of students.

- **The state needs to assure that VARAC's documentation practices are adequate and there are plans in place to run the model in the event the present system is disabled. In addition, the state should assure that there is adequate dissemination of the data.** ***(See response on pp. 55-56.)***

When conducting a process as complex as the TVAAS, it is crucial to maintain detailed and accurate directions and documentation about the process. If something were to

[43] A *t*-test is used to test sample means when the population parameter is unknown.

[44] The differences in 4th, 5th, and 8th grades are significant.

happen to the computer system at VARAC—or to key VARAC staff—Office of Education Accountability staff members are concerned that implementation of the TVAAS would have to be substantially delayed or even abandoned because there would be no way to run the model. The possibility of sabotage or accidents underscores the need to disseminate the data more widely and to have other persons, possibly in the State Department of Education, who are trained to run the model.

- **The model is not easily explainable to those affected by it.** ***(See response on p. 56.)***

Teachers, principals, and superintendents have all raised questions about the value-added assessment model because they have a great deal of difficulty understanding the process. This is understandable, given the complexity of the statistical techniques, but it makes it difficult to convince those most affected by the evaluation that it is actually an appropriate and reliable means of measuring the impact educators have on student learning.

The General Assembly deliberately crafted the statute in such a way that local school boards and educational administrators have sole discretion over how to use TVAAS classroom results in personnel evaluations. Educators have expressed a great deal of confusion and anxiety, however, over the way in which TVAAS evaluations will be used to determine issues of teacher retention and promotion. It is crucial that teachers and administrators understand the role of TVAAS in personnel evaluation. The State Department of Education and the State Board of Education need to make a concerted effort to educate teachers on the role value-added assessment is to play in holding them accountable for their performance in the classroom.

If educators do not have confidence in the value-added assessment model, they have a much greater incentive to undermine the testing and evaluation process, because the stakes of the current model for teachers are high. Such confidence could be fostered by explanations of the value-added assessment model that focus on the principles involved in the process—rather than on the statistical workings of the model—and explorations of the ways in which value-added assessment results could be used positively to help educators do a better job of helping students.

- **The "high stakes" nature of the test may create unintended incentives for both educators and students.** ***(See response on pp. 57-58.)***

A persistent warning from those persons interviewed for this evaluation was that this type of assessment program leads inevitably to the temptation to "cheat." Teachers and administrators may not be able to count on the same test being given each year, but they can easily find ways to "help" their students do well. These could include giving students breaks during the testing period or extra time to complete the test, providing students with the answers to questions when asked, or spending instructional time teaching students strategies of test-taking.[45] "Teaching to the test" is a common phenomenon in high stakes

[45]One of the superintendents interviewed for this report indicated that his elementary test scores were artificially high because elementary teachers in the district spent much of the year coaching the students on test-taking skills—robbing students of precious instructional time, and, ultimately lowering their scores in later grades. This superintendent believed that the TVAAS had shown that his second and third graders needed to lower their scores to a more realistic level. He proposed to do this by encouraging his elementary teachers to move away from the basic "drill and practice" methods they have been using—

testing, and experts agree that it can both distract from productive student learning and distort the results of standardized tests.[46]

There are statutory sanctions against those who would attempt to undermine the testing process:

> Any person found to have not followed security guidelines for administration of the TCAP test, or a successor test, including making or distributing unauthorized copies of the test, altering a grade or answer sheet, providing copies of answers or test questions, or otherwise compromising the integrity of the testing process shall be placed on immediate suspension and such actions will be grounds for dismissal, including dismissal of tenured employees. Such actions shall be grounds for revocation of state license.[47]

Sanders believes that sanctions against those compromising the integrity of the tests are sufficient to protect against widespread cheating. The value-added assessment, however, puts educators in the precarious position of administering tests that will be used to judge their performance as teachers. Even the most honest and upright teachers may find it difficult not to "teach to the test" or to administer the test in ways that will improve the scores of their students.

Another potential problem for the TVAAS is that it may offer an incentive for teachers and administrators to classify as many children as possible in special education categories. Since these students' scores are excluded by law from evaluations of teacher effect on student learning, it would benefit teachers to have both low-ability and low-achieving students classified as special education students. Low-ability students may already be classified as special education students—but low-achieving students are the ones that could benefit most from the type of information the TVAAS claims to provide. The state needs to monitor this situation carefully.[48]

Retention in grade should also be monitored. In other states with "high stakes" testing programs, the rate of retention has increased. Retention in grade is known to increase the likelihood of student drop-out later, so this practice is counterproductive in the long-term. Another major difficulty is that the test is "high stakes" for teachers, but has no consequences for students. If students understand the importance of the test for teachers, and they dislike a particular teacher, they could intentionally do poorly.[49] A small number

---

towards more innovative and creative ways of teaching. He acknowledged that such a move would undoubtedly lower test scores in his district. His chief worry was that his district might be penalized for attempting to move away from teaching to the test.

[46]See JoAnn M. Prell and Paul A. Prell, "Improving Test Scores—Teaching Test Wiseness. A Review of the Literature." ERIC Document #280900. See also Lorrie A. Shepard, "Inflated Test Score Gains: Is the Problem Old Norms or Teaching the Test?" in *Educational Measurement: Issues and Practice*, v.9 n.3, p. 15-22, Fall 1990.

[47]*TCA* §49-1-607.

[48]Many special educators argue that special education students should be tested and their progress measured as well.

[49]Apparently, this has already occurred. Personnel at the State Department of Education relate the story of a 5th grade student who was overheard attempting to organize his classmates to do poorly on the TCAP—in the hope that poor test scores would get his teacher fired.

of students could make a major (negative) difference in the scores of a particular teacher. The testing and assessment system needs to include some safeguards to prevent this type of situation from affecting a teacher's evaluation.

Recommendations

- **The TVAAS—including the model itself and the actual computer programs used to measure teacher, school, and district effect on student learning—needs to be evaluated by experts in statistics, educational measurement, and testing.**

There are important unanswered questions about the TVAAS—among them:

- Is the TVAAS appropriately applied to a norm-referenced test not constructed for this purpose?
- Why do the district and school scores vary so widely from year to year?
- What do gain scores *really* indicate? Are they pure reflections of educator influence, or are there other factors at work?
- What policy and instructional decisions can be made using the value-added data?
- What can be done to make the accountability system more comprehensible to those affected by it?

This assessment of the value-added model raises many questions about the use of the model for evaluating schools and districts. More questions will be raised when classroom evaluations become available. An evaluation of the model by both statisticians and educational measurement experts might lay to rest some of the major questions about the model and increase confidence in its use.

- **The Department of Audit should perform an Information Systems Assessment to evaluate VARAC's documentation practices and assess the safety of the testing and value-added data.**

The state needs to determine that the TVAAS is secure from accidents or potential sabotage and that policies and procedures are adequate to assure the resumption of operations in the event that these things occurred. The state also needs to ensure that adequate documentation on the value-added model has been prepared and disseminated to relevant agencies. This documentation would help to protect the state's interest in the value-added assessment program.

- **The State Board of Education and the State Department of Education need to identify unintended incentives for educators and students and consider ways to reduce their likelihood.**

The integrity of the testing process is dependent on those who take the tests and those who administer them. Teachers have a very real stake in the outcome of the testing process, and the temptation to find ways to "beat the test" will be strong. Teaching methods, or methods of test administration, may be influenced by teachers' perceptions about the best way to get high scores on the TCAP and the TVAAS—rather than by what is the most appropriate way to educate the children in their care. Anecdotal evidence indicates that teachers already spend a great deal of time preparing their students for standardized tests—that time is likely to increase as the TCAP becomes a high-stakes test for individual teachers.

If students have no stake in the testing process, they have little reason to apply themselves—yet their scores can be used in a teacher's personnel evaluation. Younger children may be eager to please their teachers by doing well on the test, but older children

may realize the importance of the tests for their teachers and respond more negatively. The ramifications of a testing system that places high stakes on teachers and no stakes on students needs to be studied.

Appendix A

Persons Interviewed

Mark Applebaum
Professor of Psychology
Coordinator, Quantitative Services Core
John F. Kennedy Center for Research on Human Development
Peabody College
Vanderbilt University

Dr. Samuel Bratton
Coordinator of Research/Evaluation
Knox County School System

Dr. Benjamin Brown
Director of Accountability Systems
State Department of Education

Dr. Paul Changas
Assistant Director
Testing and Evaluation Center
University of Tennessee at Knoxville

Wanda Copley
Executive Director
Tennessee Education Association

Dale Crews
Principal
Scotts Hill School

Dr. Sherman Dorn
Assistant Research Professor
Peabody College
Vanderbilt University

Felicia Duncan
Elementary Supervisor of Instruction
Wilson County Schools

Dr. Gary Dutton
Supervisor of Special Education and 6-12
Loudon County Schools

Dr. Thomas Fisher
Student Assessment Adminstrator
Florida Department of Education

Joel Giffin
Principal
Maryville Middle School
Maryville, Tennessee

Dr. Gene Glass
Professor, School of Education
Arizona State University

Dr. Harvey Goldstein
Institute of Education
London

Dr. Edward Haertel
Professor, School of Education
Stanford University

Dr. Gerry House
Superintendent
Memphis City Schools

Brad Hurley
Former Deputy Commissioner
State Department of Education

Dr. Richard Jaeger
Professor of Educational Research Methodology
University of North Carolina, Greensboro

Dr. Daniel Koretz
RAND Corporation

Dr. Robert Linn
Professor, School of Education
University of Colorado at Boulder

Dr. Ramon C. Littell
Professor, Department of Statistics
University of Florida, Gainesville

Ernestine McWherter
Executive Director
Tennessee Organization of School Superintendents

Charles Mitchell
Superintendent
Pickett County School System

George Nerren
Assistant Director
Tennessee School Board Association

Dr. Carol Norman
Research Division
National Education Association

Dr. Soo-Hee Park
Statistician
Testing and Evaluation Center
University of Tennessee at Knoxville

Dr. W. James Popham
IOX Associates
Culver City, California

Kip Reel
Former Deputy Commissioner
State Department of Education

Dr. William L. Sanders
Professor and Director
Value-Added Research and Assessment Center
University of Tennessee at Knoxville

Dr. Tom Satterfield
ACT

Dr. Michael Scriven
Evaluation and Development Group
Inverness, California

Dr. Lorrie Shepard
Professor, School of Education
University of Colorado at Boulder

Dr. Daniel Stufflebeam
Director, Evaluation Center and CREATE
Western Michigan University

Dr. Dan Tollett
Executive Director
Tennessee School Board Association

Karen Weeks
Research Analyst
State Board of Education

Kathy Woodall
President
Tennessee Education Association

Appendix B

References

R.D. Bingham, J.S. Herwood, and S.B. White (1991). "Evaluating Schools and Teachers Based on Student Performance," *Evaluation Review.* Vol. 15, pp.191-218.

R.D. Bingham, J.S. Herwood, and S.B. White (1992). "Another View from the Bridge: A Reply," *Evaluation Review.* Vol. 17, pp.63-68.

R. Darrell Bock, Editor (1989). *Multilevel Analysis of Educational Data*, Academic Press, Inc.

R.G. Bridge, *et al.*, (1979). *The Determinants of Educational Outcomes: The Impact of Families, Peers, Teachers, and Schools*, Ballinger Publishing Company, Cambridge, Massachusetts.

A.J. Figueredo, J. Hetherington, and L. Sechrest (1992). "Water Under the Bridge: A Response to Bingham, Herwood, and White," *Evaluation Review.* Vol. 17, pp. 40-62.

A.S. Bryk and S.W. Raudenbush, (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications.

C.R. Henderson, Jr. (1982). "Analysis of Covariance in the Mixed Model: High-Level, Nonhomogeneous, and Random Regressions," *Biometrics.* Vol.38, pp. 623-640.

Geoffrey Keppel (1982). *Design and Analysis, a Researcher's Handbook*, Second Edition. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

McLean & Sanders (1984). *Objective Component of Teacher Evaluation—A Feasibility Study, (Work Paper No. 199).* University of Tennessee, Knoxville.

McLean, Sanders, and Stroup (1991). "A Unified Approach to Mixed Linear Models," *The American Statistician.* Vol. 45, pp. 54-64.

Ned McWherter, Governor and Wayne Qualls, Commisioner of Education, (1994) *21st Century Schools Program Report Card 1994*.

Donald M. Medley, Homer Coker, and Robert S. Soar, (1984). *Measurement-Based Evaluation of Teacher Performance: An Empirical Approach*. Longman, Inc., New York.

Jason Millman and Linda Darling-Hammond, ed. (1990). *The New Handbook of Teacher Evaluation.* Sage Publications.

Melvin R. Novick, Chair (1985). *Standards for Educational and Psychological Testing*, The Committee to Develop Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Management in Education.

Ohio State Legislative Office of Education Oversight, (1993). *The Ninth-Grade Proficiency Test: Is It a Fair and Appropriate Measure of a Statewide Standard?*

W. James Popham (1988). *Educational Evaluation*, Second Edition. Prentice-Hall, Englewood Cliffs, New Jersey.

Donald Rock and others, (1991). *The Tested Achievement of the National Education Longitudinal Study of 1988 Eighth Grade Class*. ERIC Document No. 335365.

Peter H. Rossi and Howard E. Freeman (1985). *Evaluation: A systematic Approach.* Third Edition. Sage Publications.

William L. Sanders (1994). *Manual of TVAAS Seminar for Tennessee State Department of Education*, February 17-18, 1994.

William L. Sanders and Sandra Horn (1993). *An Overview of the Tennessee Value-Added Assessment System with Answers to Frequently Asked Questions.*

William L. Sanders and Sandra Horn (1994). "The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment," *Journal of Personnel Evaluation in Education* 8:299-311, 1994.

Michael Scriven (1993). *Hard-Won Lessons in Program Evaluation*. New Directions for Program Evaluation, Number 58, Summer 1993. Jossey-Bass Publishers, San Francisco.

State of Tennessee, *The Education Improvement Act of 1991*, Public Chapter No. 535, House Bill No. 752.

Tennesee State Board of Education, (1994). *Student, Teacher and School Performance 1994, Ninth Annual Report.*

Tennesee State Board of Education, (1993). *Master Plan for Tennessee Schools.*

Tennessee State Testing and Evaluation Center, (1993, 1994). *Tennessee Student Test Results* 1992-1993 & 1993-1994.

A.A. Summers, Editor, (1982). *Productivity Assessment in Education* from the series *New Directions for Testing and Measurement,* Michael Kean, Editor-in-Chief. Jossey-Bass Inc., Publishers.

Appendix C

1993-94 Report Card
Sample District—20,159 students

***Social Studies*—Estimated Means**

Grade	2	3	4	5	6	7	8
USA Norm Scale Score	652.0	691.0	713.0	735.0	745.0	749.0	761.0
1991	664.7	686.1	717.3	738.9	736.7	737.8	754.2
1992	669.1	693.9	718.4	743.9	740.8	750.5	756.7
1993	645.8	690.4	721.4	726.2	729.3	751.5	761.2
1994	642.8	674.7	705.9	733.6	739.6	747.4	754.0

This set of scores is the estimated mean scale score by grade and subject—in this case, social studies. The national norm scale score is found in the first row of scores. The school district's scores for 1991-94 are shown in the remaining rows of the table. This set of scores is used to derive the estimated gains in social studies by grade, as illustrated in the following table:

***Social Studies*—Estimated Gains**

Grade	3	4	5	6	7	8	% Norm Gain (Cumulative)
USA Norm Gain	39.0	22.0	22.0	10.0	4.0	12.0	
1992	29.3	32.1	26.5	1.9	13.5	18.9	112.0
1993	21.0	27.2	7.6	-14.5	10.6	10.5	57.3
1994	28.5	15.1	12.1	13.2	17.6	2.4	81.6

The TCAP is given to 2nd graders, but gain is not calculated until after 3rd graders take the TCAP at the end of their 3rd grade year. The national norm gains, which are calculated by VARAC, are shown in the first row of the table. This table shows the estimated gains that the sample school district students made in social studies from 1992-94.

3 Year Average	26.3R*	24.8G	15.4R*	0.2R*	13.9G	10.6R*	83.6
Standard Error	0.9	0.7	0.6	0.6	0.6	0.6	

G=Green Zone: Estimated mean gain equal to or greater than national norm.
Y=Yellow Zone: Gain below national norm by one standard error or less.
R=Red Zone: Below norm by more than one, but no more than two, standard errors.
R*=Ultra-Red: Below norm by more than two standard errors.
NG=Negative gain: No percent-of-norm calculated.

1993-94 Report Card
Sample Elementary School—292 students

***Reading*—Estimated Means**

Grade	2	3	4	5	6	7	8
USA Norm Scale Score	650.0	580.0	701.0	722.0	740.0	749.0	759.0
1991	671.2	668.4	703.6	714.6	734.7	730.5	760.4
1992	654.6	652.4	690.7	726.0	740.8	749.7	746.7
1993	658.6	684.2	743.7	732.8	744.0	763.3	771.2
1994	664.7	670.6	691.1	706.6	731.4	755.2	766.6

This set of scores is the estimated mean scale score by grade and subject—in this case, reading. The national norm scale score is found in the first row of scores. This elementary school's scores for 1991-94 are shown in the remaining rows of the table. This set of scores is used to derive the estimated gains in reading by grade, as illustrated in the following table:

***Reading*—Estimated Gains**

Grade	3	4	5	6	7	8	% Norm Gain (Cumulative)
USA Norm Gain	30.0	21.0	21.0	18.0	9.0	10.0	
1992	-18.8	22.4	22.4	26.2	15.0	16.3	76.5
1993	29.7	91.3	42.1	18.0	22.5	21.5	206.5
1994	12.0	6.9	-37.1	-1.4	11.3	3.3	NG[1]

The TCAP is given to 2nd graders, but gain is not calculated until after 3rd graders take the TCAP at the end of their 3rd grade year. The national norm gains, which are calculated by VARAC, are shown in the first row of the table. The remaining rows of the table show the estimated gains that this elementary school's students made in reading from 1992-94.

Grade	3	4	5	6	7	8	% Norm Gain (Cumulative)
3 Year Average	7.6R*	40.2G	9.1R*	14.3R	16.3G	13.7	92.8
Standard Error	4.5	3.7	3.4	3.4	3.2	3.1	

G=Green Zone: Estimated mean gain equal to or greater than national norm.
Y=Yellow Zone: Gain below national norm by one standard error or less.
R=Red Zone: Below norm by more than one, but no more than two, standard errors.
R*=Ultra-Red: Below norm by more than two standard errors.
NG=Negative gain: No percent-of-norm calculated.

[1]"Negative gain" occurs when a group of students score lower on the norm-referenced test in one year than they did on the norm-referenced test in the previous year. If the test accurately measures students' knowledge and academic skills, then negative gain means not only did students fail to learn anything in the current year, but they also forgot at least part of what they knew from the previous year.

Appendix D

1993 School System Report Card

Three-year Average (91, 92, 93) Cumulative Gain (Grades 3-8) as Percent of National Norm Gain

School System	Math	Reading	Language	Soc.	Science	Average
Anderson County	98.2%	104.5%	114.1%	100.6%	101.6%	103.8%
Clinton City	90.8%	98.9%	97.4%	94.8%	96.6%	95.7%
Oak Ridge City	109.4%	114.4%	135.4%	109.1%	107.5%	115.1%
Bedford County	102.1%	97.3%	110.5%	106.8%	104.6%	104.3%
Benton County	84.8%	97.8%	110.2%	83.1%	84.5%	92.1%
Bledsoe County	95.7%	108.7%	110.6%	104.6%	106.2%	105.2%
Blount County	88.9%	104.0%	114.4%	100.3%	95.3%	100.6%
Alcoa City	100.0%	104.1%	127.9%	93.7%	93.0%	103.7%
Maryville City	101.7%	98.2%	121.6%	81.8%	90.2%	98.7%
Bradley County	92.6%	102.0%	104.1%	88.2%	93.4%	96.0%
Cleveland City	99.6%	99.2%	124.5%	96.6%	96.9%	103.4%
Campbell County	87.0%	104.0%	109.6%	96.0%	96.2%	98.6%
Cannon County	104.0%	109.8%	118.8%	107.3%	95.4%	107.1%
Hollow Rock-Bruceton	82.9%	81.6%	82.3%	88.7%	85.1%	84.1%
Huntingdon SSD	99.4%	104.1%	116.1%	110.8%	97.6%	105.6%
McKenzie SSD	91.5%	103.0%	104.1%	86.5%	91.2%	95.3%
South Carroll Co SSD	66.6%	102.9%	95.5%	81.5%	83.9%	86.1%
West Carroll Co SSD	83.0%	118.1%	115.0%	94.9%	84.0%	99.0%
Carter County	73.3%	91.7%	93.1%	75.7%	79.6%	82.7%
Elizabethton City	101.8%	100.3%	112.7%	86.7%	80.1%	96.3%
Cheatham County	95.8%	106.9%	116.4%	96.7%	98.0%	102.8%
Chester County	92.6%	105.8%	119.0%	94.7%	101.8%	102.8%
Claiborne County	98.9%	110.9%	116.7%	96.5%	86.6%	101.9%
Clay County	82.6%	97.2%	109.6%	83.9%	84.6%	91.6%
Cocke County	89.6%	90.5%	102.8%	98.0%	84.1%	93.0%
Newport City	89.2%	109.2%	132.4%	101.4%	94.6%	105.4%
Coffee County	93.9%	110.6%	116.7%	91.0%	95.1%	101.5%
Manchester City	91.0%	96.4%	111.3%	97.5%	98.9%	99.0%
Tullahoma City	92.4%	108.9%	125.4%	98.9%	94.5%	104.0%
Crockett County	78.9%	86.7%	104.7%	75.0%	66.9%	82.4%
Alamo City	94.6%	95.2%	90.9%	95.0%	79.7%	91.1%
Bells City	61.4%	88.1%	98.2%	81.5%	74.4%	80.7%
Cumberland County	99.9%	104.4%	114.9%	99.3%	97.4%	103.2%
Davidson County	97.5%	102.0%	105.9%	97.5%	100.4%	100.7%
Decatur County	92.2%	92.0%	98.5%	85.1%	81.3%	89.8%
DeKalb County	91.5%	96.8%	92.4%	95.0%	89.4%	93.0%
Dickson County	101.0%	108.9%	111.4%	112.0%	106.8%	108.0%
Dyer County	91.6%	96.2%	116.4%	89.9%	84.3%	95.7%
Dyersburg City	98.3%	104.5%	110.8%	97.0%	106.0%	103.3%
Fayette County	107.3%	114.4%	110.3%	109.1%	101.3%	108.5%
Fentress County	94.3%	110.9%	114.4%	93.4%	97.3%	102.1%
Franklin County	94.1%	97.5%	112.0%	87.6%	85.0%	95.2%
Humboldt City	88.6%	101.3%	101.7%	86.0%	90.6%	93.7%
School System	**Math**	**Reading**	**Language**	**Soc.**	**Science**	**Average**

School System	Math	Reading	Language	Soc.	Science	Average
Milan SSD	93.4%	98.8%	111.8%	84.1%	68.3%	91.3%
Trenton SSD	104.2%	109.6%	145.1%	119.1%	102.8%	116.2%
Bradford SSD	89.3%	105.1%	104.2%	96.8%	87.1%	96.5%
Gibson County SSD	81.5%	106.2%	112.4%	95.4%	100.6%	99.2%
Giles County	99.7%	102.8%	115.0%	98.4%	96.6%	102.5%
Grainger County	92.6%	100.4%	104.7%	98.8%	91.4%	97.6%
Greene County	92.3%	108.9%	113.6%	99.3%	99.6%	102.7%
Greeneville City	94.6%	97.6%	118.1%	93.5%	105.6%	101.9%
Grundy County	86.3%	106.8%	111.9%	97.6%	88.9%	98.3%
Hamblen County	91.3%	94.6%	109.7%	94.2%	100.9%	98.1%
Hamilton County	92.7%	100.8%	117.6%	99.5%	98.7%	101.9%
Chattanooga City	93.0%	91.2%	109.7%	90.4%	96.4%	96.1%
Hancock County	87.8%	104.1%	115.3%	92.4%	107.5%	101.4%
Hardeman County	97.1%	105.0%	105.8%	99.6%	91.4%	99.8%
Hardin County	90.8%	106.3%	119.8%	92.5%	89.8%	99.9%
Hawkins County	84.0%	96.9%	100.6%	88.6%	77.9%	89.6%
Rogersville City	98.3%	99.2%	121.0%	94.6%	87.8%	100.2%
Haywood County	80.9%	86.9%	96.5%	95.3%	85.8%	89.1%
Henderson County	99.8%	110.7%	116.9%	97.8%	96.0%	104.2%
Lexington City	97.9%	95.1%	106.8%	87.6%	91.0%	95.7%
Henry County	88.3%	97.2%	109.5%	86.2%	82.7%	92.8%
Paris SSD	97.5%	94.8%	113.8%	74.8%	82.5%	92.7%
Hickman County	99.1%	103.3%	99.9%	90.1%	99.0%	98.3%
Houston County	111.3%	116.4%	126.2%	111.1%	103.0%	113.6%
Humphreys County	109.1%	104.0%	112.2%	109.5%	104.5%	107.8%
Jackson County	97.9%	107.1%	115.1%	126.5%	110.8%	111.5%
Jefferson County	94.1%	105.7%	111.7%	96.5%	94.0%	100.4%
Johnson County	96.2%	110.2%	106.6%	96.9%	107.0%	103.4%
Knox County	93.5%	106.2%	119.2%	93.9%	95.6%	101.7%
Lake County	90.3%	110.8%	122.4%	99.7%	100.9%	104.8%
Lauderdale County	108.1%	118.1%	129.3%	119.4%	106.5%	116.3%
Lawrence County	97.9%	103.0%	127.3%	96.2%	77.1%	100.3%
Lewis County	96.1%	105.1%	115.2%	105.7%	97.9%	104.0%
Lincoln County	94.4%	102.9%	99.4%	97.7%	94.6%	97.8%
Fayetteville City	96.5%	105.4%	116.3%	103.0%	95.3%	103.3%
Loudon County	90.5%	95.0%	106.8%	92.7%	95.4%	96.1%
Lenoir City	89.0%	102.1%	122.3%	101.8%	103.5%	103.8%
McMinn County	84.8%	94.9%	108.0%	90.8%	91.0%	93.9%
Athens City	87.2%	88.8%	105.6%	87.5%	88.5%	91.5%
Etowah City	62.4%	96.8%	107.7%	62.3%	43.6%	74.5%
McNairy County	89.0%	97.8%	101.6%	72.9%	70.0%	86.3%
Macon County	97.5%	105.0%	113.6%	108.8%	97.7%	104.5%
Jackson/Madison County	94.6%	102.0%	114.2%	99.5%	97.4%	101.6%
Marion County	92.6%	99.6%	100.7%	96.9%	91.0%	96.2%
Richard City SSD	126.5%	100.5%	126.0%	91.3%	97.2%	108.3%
Marshall County	94.0%	98.2%	113.2%	82.4%	86.3%	94.8%
Maury County	97.2%	104.3%	112.0%	97.2%	93.6%	100.9%
Meigs County	106.3%	108.9%	120.0%	114.2%	118.0%	113.5%

School System	Math	Reading	Language	Soc.	Science	Average
Monroe County	87.7%	99.2%	104.2%	95.6%	90.0%	95.3%
Sweetwater City	82.0%	107.4%	110.8%	107.5%	103.6%	102.3%
Montgomery County	106.2%	107.2%	117.1%	100.3%	107.4%	107.6%
Moore County	89.2%	100.3%	116.2%	103.4%	89.8%	99.8%
Morgan County	96.0%	111.5%	117.8%	93.3%	93.3%	102.4%
Obion County	85.8%	94.4%	101.7%	84.1%	99.5%	93.1%
Union City	91.0%	92.0%	104.1%	104.5%	92.4%	96.8%
Overton County	96.0%	106.8%	122.1%	110.4%	90.5%	105.2%
Perry County	93.5%	108.6%	104.3%	100.7%	86.8%	98.8%
Pickett County	95.4%	91.4%	95.7%	70.9%	75.7%	85.8%
Polk County	120.8%	123.9%	137.7%	112.7%	109.6%	120.9%
Putnam County	97.1%	100.6%	113.9%	95.0%	96.4%	100.6%
Rhea County	87.7%	115.2%	119.5%	109.0%	107.3%	107.7%
Dayton City	105.5%	107.0%	129.0%	106.0%	117.5%	113.0%
Roane County	98.8%	112.1%	114.9%	100.5%	96.8%	104.6%
Harriman City	102.8%	99.1%	107.5%	106.4%	97.3%	102.6%
Robertson County	100.6%	109.7%	121.6%	107.6%	99.6%	107.8%
Rutherford County	100.4%	107.5%	118.5%	101.1%	99.9%	105.5%
Murfreesboro City	101.4%	101.9%	111.4%	97.1%	90.8%	100.5%
Scott County	88.9%	105.5%	115.8%	110.0%	103.7%	104.8%
Oneida SSD	90.5%	101.7%	112.5%	94.7%	92.4%	98.4%
Sequatchie County	104.9%	104.1%	118.9%	97.7%	105.7%	106.3%
Sevier County	102.2%	106.0%	126.8%	98.0%	96.6%	105.9%
Shelby County	101.4%	105.4%	122.4%	93.6%	95.9%	103.7%
Memphis City	95.2%	103.1%	104.4%	103.9%	102.2%	101.8%
Smith County	114.2%	107.9%	130.9%	117.2%	104.8%	115.0%
Stewart County	99.8%	112.3%	126.4%	113.0%	111.6%	112.6%
Sullivan County	82.0%	91.1%	100.3%	76.7%	84.6%	86.9%
Bristol City	98.6%	106.0%	119.2%	93.5%	100.8%	103.6%
Kingsport City	81.3%	83.9%	95.1%	77.9%	83.1%	84.3%
Sumner County	97.6%	103.8%	118.6%	98.9%	95.5%	102.9%
Tipton County	98.3%	118.4%	121.3%	109.4%	96.2%	108.7%
Covington City	97.7%	106.6%	119.3%	110.0%	124.4%	111.6%
Trousdale County	87.3%	102.5%	107.2%	82.4%	81.3%	92.1%
Unicoi County	94.1%	110.1%	110.5%	100.8%	103.0%	103.7%
Union County	107.8%	120.6%	118.9%	100.7%	103.9%	110.4%
Van Buren County	89.4%	105.3%	125.3%	91.8%	92.0%	100.7%
Warren County	101.6%	103.2%	118.8%	100.1%	94.6%	103.6%
Washington County	90.3%	106.6%	116.2%	93.2%	90.4%	99.3%
Johnson City	82.1%	94.0%	102.9%	82.9%	99.9%	92.3%
Wayne County	95.8%	100.4%	101.8%	91.3%	88.2%	95.5%
Weakley County	98.7%	113.0%	127.6%	86.6%	88.4%	102.9%
White County	81.9%	96.8%	100.3%	98.5%	84.2%	92.3%
Williamson County	102.4%	105.6%	134.3%	104.0%	104.1%	110.1%
Franklin SSD	98.1%	95.6%	107.0%	93.0%	91.7%	97.1%
Wilson County	110.8%	112.5%	117.3%	107.8%	102.6%	110.2%
Lebanon SSD	86.1%	95.9%	103.4%	86.0%	90.1%	92.3%

Appendix E

1994 School System Report Card

Three-year Average (92, 93, 94) Cumulative Gain (Grades 3-8) as Percent of National Norm Gain

School System	Math	Reading	Language	Soc.	Science	Average
Anderson County	103.2%	113.1%	121.9%	108.1%	110.1%	111.3%
Clinton City	104.9%	110.5%	106.9%	96.9%	99.9%	103.8%
Oak Ridge	112.9%	113.0%	130.7%	104.4%	98.6%	111.9%
Bedford County	105.7%	106.2%	108.0%	100.8%	104.1%	105.0%
Benton County	94.2%	97.6%	98.1%	84.0%	78.6%	90.5%
Bledsoe County	92.4%	111.8%	124.1%	98.9%	102.6%	106.0%
Blount County	90.8%	102.7%	104.5%	96.4%	94.8%	97.8%
Alcoa City	105.0%	111.7%	135.3%	91.2%	91.2%	106.9%
Maryville City	111.4%	102.7%	125.8%	83.8%	95.4%	103.8%
Bradley County	95.6%	103.6%	94.2%	86.5%	98.8%	95.7%
Cleveland City	101.3%	104.5%	117.8%	94.4%	98.3%	103.3%
Campbell County	84.9%	96.6%	100.5%	86.5%	92.7%	92.2%
Cannon County	97.0%	113.9%	106.0%	97.6%	97.2%	102.3%
Hollow Rock-Bruceton	87.0%	90.3%	83.2%	82.7%	95.3%	87.7%
Huntingdon SSD	96.2%	105.5%	118.2%	97.4%	104.8%	104.4%
McKenzie SSD	101.4%	114.6%	117.4%	108.9%	101.5%	108.8%
South Carroll SSD	74.4%	100.4%	110.8%	94.4%	91.1%	94.2%
West Carroll SSD	85.5%	120.0%	104.8%	103.8%	99.0%	102.6%
Carter County	80.9%	90.4%	92.1%	80.4%	83.4%	85.4%
Elizabethton City	108.5%	113.6%	116.3%	92.5%	92.2%	104.6%
Cheatham County	92.2%	108.2%	106.6%	94.8%	93.1%	99.0%
Chester County	95.3%	104.1%	110.3%	91.4%	92.0%	98.6%
Claiborne County	98.6%	108.9%	110.4%	98.8%	91.4%	101.6%
Clay County	92.2%	102.8%	109.4%	90.4%	98.7%	98.7%
Cocke County	84.9%	99.3%	102.7%	83.7%	83.6%	90.9%
Newport City	96.4%	110.6%	114.9%	104.5%	113.8%	108.0%
Coffee County	99.1%	110.6%	104.9%	94.5%	101.5%	102.1%
Manchester City	95.0%	105.0%	111.5%	114.7%	102.0%	105.6%
Tullahoma City	92.8%	103.5%	110.7%	90.7%	102.1%	100.0%
Crockett County	76.3%	86.8%	89.5%	66.6%	67.4%	77.3%
Alamo City	90.4%	98.6%	85.3%	81.0%	90.6%	89.2%
Bells City	81.6%	117.4%	119.2%	91.3%	94.4%	100.8%
Cumberland County	101.0%	105.0%	104.9%	94.7%	102.7%	101.7%
Davidson County	94.6%	103.6%	97.3%	92.5%	98.4%	97.3%
Decatur County	90.1%	100.8%	105.4%	90.0%	94.2%	96.1%
DeKalb County	96.1%	103.6%	96.8%	100.3%	93.7%	98.1%
Dickson County	97.9%	104.8%	97.7%	99.4%	100.5%	100.1%
Dyer County	96.9%	101.2%	109.2%	91.8%	101.2%	100.1%
Dyersburg City	103.5%	110.0%	112.6%	102.6%	113.1%	108.4%
Fayette County	97.7%	108.2%	107.1%	107.2%	106.4%	105.3%
Fentress County	96.1%	116.1%	111.0%	94.4%	98.0%	103.1%
Franklin County	99.2%	102.0%	111.7%	93.6%	93.2%	99.9%
Humboldt City	92.2%	103.2%	103.4%	87.5%	98.8%	97.0%
School System	**Math**	**Reading**	**Language**	**Soc.**	**Science**	**Average**

School System	Math	Reading	Language	Soc.	Science	Average
Milan SSD	99.8%	95.7%	104.0%	80.6%	78.7%	91.8%
Trenton SSD	105.3%	110.0%	123.5%	94.9%	106.3%	108.0%
Bradford SSD	88.0%	103.1%	89.7%	88.8%	103.9%	94.7%
Gibson SSD	94.0%	106.7%	109.7%	96.1%	108.0%	102.9%
Giles County	99.8%	102.4%	105.0%	95.4%	99.0%	100.3%
Grainger County	87.6%	102.4%	85.6%	86.4%	96.5%	91.7%
Greene County	93.4%	115.3%	108.5%	98.8%	104.3%	104.1%
Greeneville City	90.4%	100.4%	92.0%	82.7%	83.7%	89.9%
Grundy County	89.0%	106.2%	98.2%	95.5%	90.8%	95.9%
Hamblen County	93.7%	103.8%	102.7%	94.7%	109.2%	100.8%
Hamilton County	97.0%	103.8%	113.0%	95.5%	101.3%	102.1%
Chattanooga City	89.7%	92.4%	95.8%	83.6%	88.0%	89.9%
Hancock County	93.0%	109.8%	117.1%	97.7%	113.9%	106.3%
Hardeman County	92.6%	99.7%	89.7%	82.4%	99.4%	92.8%
Hardin County	100.5%	123.7%	128.6%	101.3%	100.2%	110.9%
Hawkins County	85.6%	96.6%	96.7%	85.5%	87.8%	90.4%
Rogersville City	97.3%	95.1%	109.0%	84.1%	83.4%	93.8%
Haywood County	88.7%	96.9%	92.6%	91.3%	102.9%	94.5%
Henderson County	111.4%	116.2%	116.3%	104.3%	112.9%	112.2%
Lexington City	99.2%	99.4%	103.4%	84.4%	102.5%	97.8%
Henry County	90.6%	99.6%	97.4%	89.8%	91.7%	93.8%
Paris SSD	87.3%	91.7%	103.4%	69.0%	82.4%	86.8%
Hickman County	100.1%	101.6%	91.7%	93.1%	103.0%	97.9%
Houston County	98.3%	110.1%	113.4%	102.1%	105.6%	105.9%
Humphreys County	111.4%	103.8%	107.3%	108.2%	112.8%	108.7%
Jackson County	104.7%	108.1%	115.0%	118.9%	111.9%	111.7%
Jefferson County	96.5%	108.0%	105.0%	101.2%	100.3%	102.2%
Johnson County	95.8%	114.0%	101.1%	94.9%	102.8%	101.7%
Knox County	97.9%	110.7%	116.4%	95.5%	102.2%	104.5%
Lake County	87.4%	113.2%	110.0%	89.7%	87.4%	97.5%
Lauderdale County	110.9%	121.7%	124.0%	108.5%	116.5%	116.3%
Lawrence County	98.4%	108.1%	120.4%	100.5%	91.5%	103.8%
Lewis County	105.2%	108.7%	115.6%	97.8%	101.7%	105.8%
Lincoln County	95.2%	102.5%	96.2%	95.3%	102.3%	98.3%
Fayetteville City	87.9%	100.4%	121.1%	105.1%	86.1%	100.1%
Loudon County	83.0%	100.4%	102.0%	83.8%	92.5%	92.4%
Lenoir City	88.0%	94.1%	117.9%	95.1%	93.4%	97.7%
McMinn County	93.0%	97.4%	109.7%	88.5%	95.5%	96.8%
Athens City	88.7%	90.9%	104.2%	80.9%	89.5%	90.8%
Etowah City	70.5%	93.8%	95.3%	86.0%	82.3%	85.6%
McNairy County	96.0%	108.1%	106.4%	88.5%	89.8%	97.7%
Macon County	97.5%	104.7%	111.1%	98.4%	104.6%	103.3%
Jackson/Madison County	91.6%	96.2%	97.7%	88.8%	94.4%	93.7%
Marion County	91.0%	104.2%	99.7%	93.0%	97.8%	97.1%
Richard City SSD	127.3%	96.3%	112.1%	90.7%	84.1%	102.1%
Marshall County	100.3%	95.3%	108.5%	96.8%	104.9%	101.2%
Maury County	100.0%	101.8%	100.8%	96.4%	95.4%	98.8%
Meigs County	91.6%	114.1%	115.9%	87.6%	107.9%	103.4%

School System	Math	Reading	Language	Soc.	Science	Average
Monroe County	87.9%	103.8%	92.7%	95.6%	99.4%	95.9%
Sweetwater City	79.8%	96.9%	111.8%	102.6%	92.1%	96.7%
Montgomery County	107.8%	114.0%	114.8%	98.5%	107.6%	108.6%
Moore County	91.1%	100.4%	98.3%	95.5%	102.1%	97.5%
Morgan County	93.6%	107.3%	89.3%	81.8%	82.6%	90.9%
Obion County	86.8%	92.0%	91.2%	84.5%	97.0%	90.3%
Union City	95.7%	99.7%	99.1%	95.4%	106.3%	99.2%
Overton County	92.7%	103.9%	113.3%	99.4%	92.5%	100.4%
Perry County	95.9%	111.6%	98.8%	94.8%	100.9%	100.4%
Pickett County	94.9%	85.2%	97.7%	74.2%	86.2%	87.6%
Polk County	126.7%	119.6%	131.3%	113.6%	118.4%	121.9%
Putnam County	103.3%	105.7%	108.8%	88.5%	99.3%	101.1%
Rhea County	92.7%	109.5%	110.6%	106.5%	111.9%	106.3%
Dayton City	94.9%	105.5%	110.3%	97.6%	102.0%	102.1%
Roane County	108.4%	118.1%	117.5%	104.3%	107.1%	111.1%
Harriman City	110.6%	111.2%	112.5%	104.2%	103.8%	108.5%
Robertson County	95.0%	108.0%	105.9%	92.7%	98.1%	99.9%
Rutherford County	100.4%	108.4%	109.6%	94.3%	98.1%	102.2%
Murfreesboro City	96.1%	100.6%	97.7%	84.5%	87.9%	93.4%
Scott County	89.8%	99.1%	101.0%	88.0%	96.1%	94.8%
Oneida SSD	96.9%	85.1%	107.6%	88.2%	90.0%	93.6%
Sequatchie County	105.5%	107.1%	112.1%	90.4%	111.2%	105.3%
Sevier County	101.2%	103.8%	121.7%	95.1%	97.3%	103.8%
Shelby County	97.9%	110.3%	118.1%	94.4%	97.9%	103.7%
Memphis City	84.0%	90.4%	84.2%	85.6%	98.5%	88.6%
Smith County	116.6%	107.7%	117.2%	100.7%	108.0%	110.0%
Stewart County	98.9%	120.5%	127.9%	111.6%	115.1%	114.8%
Sullivan County	81.8%	94.6%	90.7%	76.5%	86.9%	86.1%
Bristol City	95.0%	106.0%	113.6%	87.5%	95.2%	99.5%
Kingsport City	87.1%	92.2%	92.2%	78.8%	89.7%	88.0%
Sumner County	98.4%	106.3%	109.7%	96.3%	97.8%	101.7%
Tipton County	92.9%	108.2%	103.9%	92.2%	95.5%	98.6%
Covington City	93.9%	92.6%	105.9%	88.3%	122.4%	100.6%
Trousdale County	87.8%	97.1%	96.0%	73.4%	84.0%	87.7%
Unicoi County	89.5%	109.3%	94.9%	98.1%	112.6%	100.9%
Union County	110.6%	121.4%	116.1%	111.1%	112.2%	114.3%
Van Buren County	80.0%	103.4%	112.1%	90.4%	88.4%	94.9%
Warren County	109.5%	112.6%	123.3%	102.3%	99.3%	109.4%
Washington County	95.3%	111.6%	110.4%	87.1%	91.9%	99.3%
Johnson City	89.3%	98.4%	98.9%	89.3%	97.3%	94.6%
Wayne County	94.6%	102.6%	95.0%	88.3%	88.5%	93.8%
Weakley County	97.2%	113.4%	119.9%	82.9%	92.6%	101.2%
White County	89.2%	103.3%	105.3%	91.2%	92.0%	96.2%
Williamson County	105.7%	106.7%	119.9%	98.4%	103.8%	106.9%
Franklin SSD	94.0%	91.8%	96.0%	85.2%	82.8%	89.9%
Wilson County	108.2%	111.8%	106.3%	101.4%	98.8%	105.3%
Lebanon SSD	86.5%	97.5%	93.4%	82.4%	79.9%	87.9%

Appendix F

Changes from 1993 to 1994 Report Card

Three-year Average Cumulative Gain (Grades 3-8) as Percent of National Norm Gain

School System	Math Change	Reading Change	Language Change	Soc. Change	Science Change	Average Change
Anderson County	5.1%	8.6%	7.8%	7.5%	8.4%	7.5%
Clinton City	14.0%	11.6%	9.5%	2.1%	3.3%	8.1%
Oak Ridge	3.6%	-1.4%	-4.7%	-4.7%	-8.9%	-3.2%
Bedford County	3.7%	8.8%	-2.5%	-6.0%	-0.5%	0.7%
Benton County	9.5%	-0.2%	-12.1%	0.9%	-5.9%	-1.6%
Bledsoe County	-3.3%	3.1%	13.5%	-5.6%	-3.6%	0.8%
Blount County	1.8%	-1.4%	-10.0%	-3.9%	-0.5%	-2.8%
Alcoa City	5.0%	7.5%	7.4%	-2.5%	-1.9%	3.1%
Maryville City	9.7%	4.6%	4.2%	2.0%	5.2%	5.1%
Bradley County	3.0%	1.6%	-9.8%	-1.8%	5.4%	-0.3%
Cleveland City	1.7%	5.4%	-6.7%	-2.2%	1.5%	-0.0%
Campbell County	-2.1%	-7.4%	-9.1%	-9.5%	-3.5%	-6.3%
Cannon County	-7.0%	4.1%	-12.8%	-9.7%	1.7%	-4.7%
Hollow Rock-Bruceton	4.1%	8.7%	0.9%	-5.9%	10.2%	3.6%
Huntingdon SSD	-3.2%	1.4%	2.1%	-13.3%	7.2%	-1.2%
McKenzie SSD	9.9%	11.6%	13.3%	22.4%	10.3%	13.5%
South Carroll SSD	7.8%	-2.6%	15.3%	12.8%	7.1%	8.1%
West Carroll SSD	2.5%	2.0%	-10.1%	8.9%	15.0%	3.6%
Carter County	7.6%	-1.4%	-1.0%	4.6%	3.7%	2.7%
Elizabethton City	6.6%	13.3%	3.6%	5.7%	12.1%	8.3%
Cheatham County	-3.6%	1.3%	-9.8%	-1.9%	-4.8%	-3.8%
Chester County	2.7%	-1.8%	-8.7%	-3.3%	-9.8%	-4.2%
Claiborne County	-0.3%	-2.0%	-6.2%	2.3%	4.7%	-0.3%
Clay County	9.6%	5.6%	-0.2%	6.5%	14.1%	7.1%
Cocke County	-4.7%	8.8%	-0.0%	-14.4%	-0.4%	-2.1%
Newport City	7.2%	1.3%	-17.6%	3.1%	19.2%	2.7%
Coffee County	5.2%	0.0%	-11.8%	3.5%	6.4%	0.6%
Manchester City	3.9%	8.6%	0.2%	17.2%	3.2%	6.6%
Tullahoma City	0.4%	-5.4%	-14.7%	-8.2%	7.6%	-4.1%
Crockett County	-2.5%	0.0%	-15.2%	-8.4%	0.6%	-5.1%
Alamo City	-4.2%	3.4%	-5.6%	-14.0%	10.8%	-1.9%
Bells City	20.1%	29.3%	21.0%	9.9%	19.9%	20.0%
Cumberland County	1.1%	0.6%	-10.0%	-4.6%	5.3%	-1.5%
Davidson County	-2.9%	1.6%	-8.6%	-5.0%	-2.0%	-3.4%
Decatur County	-2.1%	8.8%	6.9%	4.9%	12.9%	6.3%
DeKalb County	4.5%	6.7%	4.4%	5.3%	4.3%	5.1%
Dickson County	-3.1%	-4.1%	-13.7%	-12.6%	-6.3%	-8.0%
Dyer County	5.3%	5.0%	-7.2%	1.9%	16.9%	4.4%
Dyersburg City	5.2%	5.5%	1.9%	5.6%	7.1%	5.1%
Fayette County	-9.6%	-6.2%	-3.2%	-1.9%	5.1%	-3.1%
Fentress County	1.9%	5.2%	-3.4%	1.0%	0.6%	1.1%
Franklin County	5.1%	4.4%	-0.2%	6.0%	8.2%	4.7%

School System	Math Change	Reading Change	Language Change	Soc. Change	Science Change	Average Change
Humboldt City	3.6%	1.9%	1.7%	1.5%	8.2%	3.4%
Milan SSD	6.4%	-3.1%	-7.8%	-3.5%	10.4%	0.5%
Trenton SSD	1.1%	0.4%	-21.6%	-24.2%	3.5%	-8.2%
Bradford SSD	-1.3%	-2.0%	-14.5%	-8.0%	16.8%	-1.8%
Gibson SSD	12.5%	0.5%	-2.8%	0.7%	7.4%	3.7%
Giles County	0.1%	-0.4%	-10.0%	-3.0%	2.4%	-2.2%
Grainger County	-5.1%	2.0%	-19.1%	-12.4%	5.0%	-5.9%
Greene County	1.1%	6.4%	-5.0%	-0.5%	4.8%	1.3%
Greeneville City	-4.2%	2.9%	-26.1%	-10.8%	-21.9%	-12.0%
Grundy County	2.7%	-0.6%	-13.7%	-2.1%	1.9%	-2.4%
Hamblen County	2.3%	9.3%	-7.0%	0.4%	8.3%	2.7%
Hamilton County	4.3%	3.0%	-4.5%	-4.0%	2.6%	0.3%
Chattanooga City	-3.3%	1.3%	-13.8%	-6.7%	-8.4%	-6.2%
Hancock County	5.2%	5.6%	1.9%	5.3%	6.3%	4.9%
Hardeman County	-4.5%	-5.3%	-16.1%	-17.1%	8.0%	-7.0%
Hardin County	9.7%	17.4%	8.7%	8.9%	10.4%	11.0%
Hawkins County	1.6%	-0.3%	-3.8%	-3.1%	9.9%	0.9%
Rogersville City	-1.0%	-4.1%	-12.0%	-10.4%	-4.3%	-6.4%
Haywood County	7.9%	10.0%	-3.8%	-4.0%	17.1%	5.4%
Henderson County	11.5%	5.5%	-0.6%	6.5%	16.9%	7.9%
Lexington City	1.3%	4.3%	-3.4%	-3.2%	11.5%	2.1%
Henry County	2.2%	2.3%	-12.1%	3.6%	9.0%	1.0%
Paris SSD	-10.3%	-3.1%	-10.4%	-5.8%	-0.0%	-5.9%
Hickman County	1.0%	-1.7%	-8.2%	2.9%	3.9%	-0.4%
Houston County	-13.0%	-6.3%	-12.8%	-9.0%	2.7%	-7.7%
Humphreys County	2.4%	-0.2%	-4.9%	-1.4%	8.3%	0.9%
Jackson County	6.8%	1.0%	-0.2%	-7.6%	1.2%	0.2%
Jefferson County	2.4%	2.4%	-6.7%	4.7%	6.4%	1.8%
Johnson County	-0.3%	3.7%	-5.5%	-2.1%	-4.2%	-1.7%
Knox County	4.4%	4.5%	-2.8%	1.6%	6.5%	2.9%
Lake County	-2.9%	2.4%	-12.4%	-10.0%	-13.5%	-7.3%
Lauderdale County	2.8%	3.6%	-5.3%	-10.9%	10.1%	0.0%
Lawrence County	0.4%	5.1%	-6.9%	4.3%	14.4%	3.5%
Lewis County	9.1%	3.6%	0.5%	-7.9%	3.8%	1.8%
Lincoln County	0.8%	-0.3%	-3.2%	-2.5%	7.6%	0.5%
Fayetteville City	-8.7%	-5.0%	4.7%	2.1%	-9.2%	-3.2%
Loudon County	-7.6%	5.4%	-4.8%	-8.9%	-2.9%	-3.7%
Lenoir City	-1.0%	-7.9%	-4.4%	-6.7%	-10.1%	-6.0%
McMinn County	8.3%	2.5%	1.7%	-2.3%	4.5%	2.9%
Athens City	1.6%	2.1%	-1.4%	-6.6%	0.9%	-0.7%
Etowah City	8.1%	-3.0%	-12.4%	23.7%	38.7%	11.0%
McNairy County	7.0%	10.3%	4.8%	15.6%	19.8%	11.5%
Macon County	0.0%	-0.2%	-2.5%	-10.5%	6.9%	-1.3%
Jackson/Madison County	-3.0%	-5.8%	-16.5%	-10.8%	-3.0%	-7.8%
Marion County	-1.7%	4.6%	-1.0%	-3.9%	6.8%	0.9%
Richard City SSD	0.8%	-4.1%	-13.9%	-0.6%	-13.1%	-6.2%
Marshall County	6.3%	-2.9%	-4.7%	14.4%	18.6%	6.3%

School System	Math Change	Reading Change	Language Change	Soc. Change	Science Change	Average Change
Maury County	2.7%	-2.5%	-11.3%	-0.8%	1.7%	-2.0%
Meigs County	-14.7%	5.2%	-4.2%	-26.6%	-10.1%	-10.1%
Monroe County	0.2%	4.6%	-11.5%	0.0%	9.4%	0.6%
Sweetwater City	-2.2%	-10.5%	1.0%	-4.9%	-11.5%	-5.6%
Montgomery County	1.6%	6.8%	-2.3%	-1.8%	0.2%	0.9%
Moore County	1.9%	0.0%	-17.9%	-7.9%	12.3%	-2.3%
Morgan County	-2.4%	-4.2%	-28.5%	-11.5%	-10.8%	-11.5%
Obion County	1.0%	-2.5%	-10.5%	0.3%	-2.5%	-2.8%
Union City	4.6%	7.7%	-5.0%	-9.0%	13.9%	2.4%
Overton County	-3.3%	-3.0%	-8.7%	-11.0%	2.0%	-4.8%
Perry County	2.4%	3.0%	-5.5%	-5.9%	14.1%	1.6%
Pickett County	-0.6%	-6.2%	2.0%	3.2%	10.5%	1.8%
Polk County	6.0%	-4.3%	-6.4%	1.0%	8.8%	1.0%
Putnam County	6.3%	5.1%	-5.2%	-6.5%	2.9%	0.5%
Rhea County	5.0%	-5.7%	-8.9%	-2.5%	4.6%	-1.5%
Dayton City	-10.5%	-1.5%	-18.7%	-8.4%	-15.6%	-10.9%
Roane County	9.6%	6.0%	2.6%	3.8%	10.3%	6.4%
Harriman City	7.8%	12.1%	5.0%	-2.1%	6.5%	5.8%
Robertson County	-5.5%	-1.7%	-15.7%	-14.9%	-1.5%	-7.9%
Rutherford County	0.0%	0.9%	-8.9%	-6.7%	-1.8%	-3.3%
Murfreesboro City	-5.3%	-1.4%	-13.7%	-12.5%	-2.9%	-7.1%
Scott County	0.9%	-6.4%	-14.7%	-22.0%	-7.6%	-10.0%
Oneida SSD	6.4%	-16.5%	-4.9%	-6.6%	-2.4%	-4.8%
Sequatchie County	0.6%	3.0%	-6.8%	-7.3%	5.5%	-1.0%
Sevier County	-1.0%	-2.2%	-5.1%	-2.9%	0.7%	-2.1%
Shelby County	-3.5%	4.9%	-4.3%	0.9%	2.0%	0.0%
Memphis City	-11.2%	-12.7%	-20.1%	-18.3%	-3.7%	-13.2%
Smith County	2.4%	-0.2%	-13.7%	-16.5%	3.1%	-5.0%
Stewart County	-0.9%	8.2%	1.5%	-1.4%	3.5%	2.2%
Sullivan County	-0.2%	3.5%	-9.6%	-0.2%	2.3%	-0.8%
Bristol City	-3.6%	0.0%	-5.7%	-6.0%	-5.6%	-4.2%
Kingsport City	5.8%	8.3%	-2.9%	0.9%	6.6%	3.7%
Sumner County	0.8%	2.5%	-8.9%	-2.6%	2.3%	-1.2%
Tipton County	-5.4%	-10.2%	-17.3%	-17.1%	-0.8%	-10.2%
Covington City	-3.8%	-14.0%	-13.4%	-21.7%	-2.0%	-11.0%
Trousdale County	0.5%	-5.3%	-11.2%	-9.0%	2.7%	-4.5%
Unicoi County	-4.5%	-0.8%	-15.6%	-2.6%	9.6%	-2.8%
Union County	2.8%	0.9%	-2.8%	10.5%	8.3%	3.9%
Van Buren County	-9.5%	-1.9%	-13.2%	-1.3%	-3.6%	-5.9%
Warren County	7.9%	9.4%	4.5%	2.3%	4.7%	5.7%
Washington County	5.1%	5.0%	-5.8%	-6.1%	1.6%	-0.0%
Johnson City	7.2%	4.4%	-4.0%	6.5%	-2.6%	2.3%
Wayne County	-1.3%	2.2%	-6.8%	-3.0%	0.4%	-1.7%
Weakley County	-1.5%	0.4%	-7.7%	-3.7%	4.2%	-1.6%
White County	7.3%	6.6%	5.0%	-7.2%	7.9%	3.9%
Williamson County	3.3%	1.0%	-14.4%	-5.6%	-0.3%	-3.2%
Franklin SSD	-4.2%	-3.8%	-11.0%	-7.9%	-8.9%	-7.1%

School System	Math Change	Reading Change	Language Change	Soc. Change	Science Change	Average Change
Wilson County	-2.7%	-0.7%	-11.0%	-6.5%	-3.7%	-4.9%
Lebanon SSD	0.4%	1.6%	-10.0%	-3.7%	-10.2%	-4.4%

Appendix G
Response by Dr. William L. Sanders

The following remarks have been condensed from a written response received by the Office of Education Accountability on April 5, 1995. As much as possible, Dr. Sanders' complete original text has been incorporated, but his detailed separate responses to the executive summary and the body of the report were condensed. Some of the original text has been eliminated because it referred to portions of the draft of the report that were changed or eliminated in the final version.

Need For Independent Evaluation

(Executive Summary, p. i; Report Text, p. 9)

An independent evaluation could address a number of issues related to the value-added assessment. Any such evaluation should involve non-biased persons with experience in statistical mixed modeling, education testing, and assessment. However, the report should include information concerning the tremendous amount of validation that has already transpired.

As has been stated many times, a competent, objective, and independent review is welcomed. Even though a "comprehensive" outside review of TVAAS has not been completed as of yet, considerable evaluation and validation has been completed in many different ways at many different levels over the past 13 years. A report such as this with a high probability of widespread distribution should contain, at a minimum, acknowledgment of the previous and current validation efforts. These efforts will be listed by categories.

I. Mixed-Model Theory and Methods

A whole literature exists built on the original work of C. R. Henderson, the Cornell animal geneticist, who originated the concept of BLUP. However, similar developments have evolved from other disciplines including the hierarchical linear modeling (HLM) techniques that are being espoused and used by some educational measurement specialists such as Harvey Goldstein in Great Britain. Kalman filtering from the engineering sciences is based upon similar concepts. David Harville, Department of Statistics, Iowa State University, has authored publications that detail the relationship among these various techniques and their relationship to empirical Bayesian statistical theory. Certainly, the underpinning of TVAAS is supported by rigorous statistical theory. Additionally, the TVAAS application has been presented at professional statistical meetings at the local, regional and national levels without receiving the first challenge to the statistical validity of the application.

II. Empirical Validation from the Original Three Pilot Studies and Dr. Casteel's Dissertation.

A. Knox County Study

In 1983, after the analysis of the data had been completed, including the estimation of the teacher effects, an invitation was

extended by Knox County System administrators to learn of the results. In attendance at that meeting were Dr. Sarah Simpson, Supervisor of Instruction, Dr. Sam Bratton, Supervisor of Testing and Evaluation, and other supervisors. At that meeting the first validation exercise was conducted. The names of teachers, one at a time, were read from the top or bottom ten teachers from the profiles for each of grades 3, 4, and 5 and the group was asked to guess from which end of the distribution each name came. The agreement between the group's opinion of where teachers would appear on the distribution and the actual findings of the study was over 90%. Any of those present could be contacted to verify the results of the exercise.

B. Blount County Study

The Blount County Study was conducted as part of the dissertation of David Cook. Dr. Cook continues to work in the Blount County System. As part of his research objectives, David asked each of the principals to forecast whether or not each teacher (fifth grade) would profile in the top, mid or bottom third of the teachers in Blount County. The principals forecasted most of the bottom 1/3 teachers, were reasonably accurate in their forecasts of the top math teachers, but did not distinguish between the top and mid reading and language arts teachers.

Also, data for student ability (IQ scores) were available in this study. Much attention was directed to the need for an independent measure of student ability during the discussions about TVAAS on EDPOLYAN. The results from the Blount County study were virtually identical to the Knox County results in which the IQ scores were not available. The similarities of the results strongly suggest that the inclusion of IQ scores did not contribute any additional information and was not necessary to insure that differences in student ability were adequately accounted.

C. Chattanooga City Study

The Chattanooga City study was conducted as part of the dissertation of John Kellifer. Dr. Kellifer continues to work in the Chattanooga City System. Dr. Kellifer's research question dealt primarily with attempting to measure the relationship between other school evaluation models and the value-added assessment results from student achievement data. He found virtually no relationship between these other models for the evaluation of schools, which were heavily loaded with faculty attitudinal measures, and the value-added profiles of the Chattanooga City Schools.

Also from the Chattanooga data, the value-added profiles of schools were found to have no relationship with the racial composition of schools, a finding that has been repeatedly confirmed from the statewide data. Additionally, the distribution of

the teacher effects were very similar to the Knox and Blount County distributions.

D. Dr. DiAnn Casteel's Dissertation

Dr. Casteel completed her dissertation at East Tennessee State University in 1994. She attempted to relate the TVAAS school value-added estimates to characteristics of the schools in the first congressional district of Tennessee. She confirmed, as was found in the earlier studies, that the location of the building was not a predictor of the effectiveness of the schools as measured by TVAAS.

III. Software Validation

How are the results coming from the TVAAS software documented for accuracy?

The large number of mixed-model equations required for all but the smallest Tennessee school system makes their solution, utilizing commercially available mixed-model software, impossible even on the largest mainframe computer at UT-K. Therefore, it has been necessary to develop original software to complete the computing tasks required by TVAAS.

The "black box" is the software which sets up and solves the mixed-model equations; to verify the computational accuracy, two different approaches to verification have been deployed. For the first approach, a very small problem was analyzed by commercially available software and with the "black box" and the results compared. Complete agreement was found.

The second form of validation required the construction of the "gray box." The "gray box" is software that has been built using a standard computer matrix language; it can accommodate a modest sized problem. Because it can handle a larger data set than the commercial software, it offers a better test to check the accuracy of the "black box" results. Again the results were found to be identical. The results of these confirming tests are available at UT-VARAC.

Additionally, three nationally recognized statistical mixed-model experts have been asked to construct data sets (independently of each other); obtain solutions using software of their choosing; submit the data for processing through the TVAAS software, under their supervision; compare the results; and report their findings. Also, these experts are being asked to review the empirical evidence that the statistical assumptions of the model are being met. The reports of their finding will be available within the next several weeks.

How Can the Accuracy of the Merging of the Data Be Checked?

Since each school system has a copy of each child's TCAP report, random samples of students could be drawn to check the accuracy of the merging step.

Ready Access to TVAAS Data and Software

It has been publicly stated numerous times that cooperative research with investigators at the university, system, and school level is encouraged and welcome. The data, facilities, and expertise of the UT-VARAC center are available to assist anyone who has a research hypothesis that can be evaluated with the total resources of the center. Several individuals have availed themselves of this opportunity. A partial list includes the following:

Dr. David Burrell, former graduate student, ETSU;
Dr. DiAnn Casteel, former graduate student, ETSU;
Dr. Sam Bratton, form Knoxville, and others from Nashville Metro and Memphis City, working on a project to measure the impact of multiple school changes within a year on student performance;
Dr. Barbara Nye, studying the impact of the science teaching initiative at Tennessee State University;
Dr. Peggy Harris, to evaluate the impact of NDN programs;
Dr. Russell French, project leader of a cooperative effort between the UTK College of Education, the State Testing and Evaluation Center, and UT-VARAC, to determine ways to assist systems in minimizing the building change effect and to cope with intraclassroom diversity.

Thus, to leave the impression with the readers of this report that these resources have been hoarded is totally without basis.

As to the use of the software for validation, several points need to be made. 1) The hardware requirement necessary to fit the data via the mixed-model equations is huge. The dedicated RS-6000 workstation has 1 gigabyte of RAM. 2) The software as of yet is not a deliverable product as per the contractual responsibility. 3) The accuracy of the software can be validated on site by independent knowledgeable persons by inspecting the validation studies that are routinely conducted or by developing their own data set off site, allowing us to process that data using TVAAS software, and then comparing our results to theirs. In fact, such a process is currently under way as was previously mentioned.

The specific mathematical model for schools was posted on the Internet bulletin board EDPOLYAN. This is in addition to presentations at professional meetings and in the Sanders and Horn 1994 paper, which was cited in your bibliography.

Office of Education Accountability Comment

Although Sanders has been working on the value-added assessment model since the early 1980s, he has not submitted the model to the wider professional community for validation. In fact, several of the nationally-recognized educational measurement, testing, and statistical experts interviewed for this report had never heard of the TVAAS. Others were familiar with the concept but had little or no information on which to base an opinion of the model itself.

Office of Education Accountability analysts believe that validation by the educational measurement and testing communities is a necessary step in the evaluation of the TVAAS. The assumptions behind the value-added assessment model and the results it has generated are important enough—and controversial enough— to be debated in the nation's major statistical and educational measurement journals. The "tremendous amount of validation that has already transpired" consists largely of unpublished dissertations and conference papers. These are rarely subjected to the same level of scrutiny by acknowledged experts as are journal articles and replication efforts.

Unexplained Variability in National Norms

(See Executive Summary, p. i-ii ; Report text, pp. 10-11)

Although national norm gains do vary from year to year, the analysis of the data demonstrates that Tennessee student gains consistently parallel national norm gains in most subjects and grades. In science, where the difference is most pronounced, the impressive statewide gains made by Tennessee students in the 4th grade are not repeated in the 5th grade. Measured over the two academic years, however, the total gains by Tennessee students are almost identical to the two-year national norm gains. Over six years, Tennessee science gains vary by a maximum of only 4.5 percent.

This issue speaks to an important purpose of the value-added assessment. Long term improvement in student learning requires that efforts to strengthen a school's curricula and instruction not be limited to a single subject or grade. For example, if student gains are strong in fourth grade science, to achieve similar gains in the 5th grade might require that 4th grade curricula not be repeated.

The issue of expected gains in 7th grade social studies warrants further examination. More than any other subject area, the curricula for 7th grade social studies differ throughout the country. These differences would not statistically affect the ability to compare gains among systems and schools, but they would have a potential effect on comparisons at grade level.

It should be also noted that teacher effects are estimated around the mean for each system respectively. Thus, unbiased estimates of the teacher effects are not directly dependent upon the gains from the national norm curves. Also, it should be noted that the Commissioner recommended and the State Board of Education approved a set of standards for use in evaluating compliance with the EIA that is primarily based upon CUMULATIVE gain, which should mitigated some of the concerns expressed in this section.

<u>Office of Education Accountability Response</u>

In his written response, Sanders did not address the question raised in the report: are national norm gains the most appropriate benchmark by which to judge Tennessee educators? The following points remain issues of concern:

- TCAP tests were not built for the purposes of the TVAAS, so this raises the question of the appropriateness of basing an accountability system on them. CTB/McGraw Hill has assured the state that the match between the tests and

the curriculum is sufficient for the purposes of the TVAAS, but this has not been independently confirmed.

- Overall cumulative gains for Tennessee students in grades 3-8 are similar to national norm gains—but the wide variety of grade configurations in Tennessee schools means that many schools are measured only against the national norm gains in the few grades that reflect their particular grade configuration. This makes it impossible to avoid focusing on gains in specific grades.
- There is absolutely no evidence to support Sanders' contention that gains are affected because curricula is being repeated—particularly in social studies and science. The state has a general curriculum framework and each district has a more detailed curriculum guideline for each grade in each subject. There is little or no chance that 4th grade teachers would be teaching their students a 5th grade science curriculum—or that 5th grade teachers would repeat the 4th grade science curriculum.
- The chance that curricula might be repeated in reading, language arts, and math is more likely—but there is no data to show that this is occurring. Office of Education Accountability analysts feel that more evidence is needed before this hypothesis can be used in an explanatory way.
- If Sanders feels that "extracting the scores for an individual subject in an individual grade distorts the meaning and purpose of the value-added assessment," this raises questions about using the model to evaluate teachers or adjust instruction.
- The Office of Education Accountability is still concerned that the great variation in national norm gain expectations from grade to grade results in evaluations that appear more favorable to Tennessee middle schools than to the state's elementary schools and their teachers.

Large Changes That Are Unexplained

(See Executive Summary, p. ii; Report text, pp.11-14)

This critique would be valid if one believed learning could not be improved in Tennessee's schools. Value-added scores do vary significantly from year to year within many systems and within many schools. Indeed, the very purpose of the value-added assessment would be suspect if scores did not change. If academic gains did not change, it would suggest that all Tennessee students are performing at maximum level, or that teachers, administrators and students have no response whatsoever either to deficiencies or accomplishments reflected by previous scores.

A lack of response in fact did occur during the pilot testing of the value-added instrument in the Knox, Blount and Chattanooga City schools because those studies were done retrospectively with no feedback to the schools...Predictably, when the value-added instrument became formalized statewide in 1992, many teachers and administrators became much more conscious of trying to address weaknesses and build upon strengths identified by the test scores.

The concerns that led to the broad reforms contained in the 21st Century Schools programs were based in large measure on the assumption that there were deficiencies,

including wide gaps in how much students were learning, among Tennessee's 139 systems and 1,600 schools. The documented presence of such disparities in academic performance should not come as a surprise. Rather, it should reinforce the belief expressed by the Legislature that it is possible in each grade and each subject to improve how much our students learn.

Actually, such changes in variation are to be expected at this stage of implementation of TVAAS and were, in fact, predicted before TVAAS was implemented. It may be noted that the variation has increased as TVAAS scores have become available to Tennessee schools and systems. Although the actual reasons for this variation will be discussed later in this section, some questions that may have presented themselves to the casual observer as possible explanations for these changes will be addressed.

A. Are there severe inconsistencies in the tests among years?

No. A statewide analysis of the distribution of test scores indicates that these tests are equivalent with regard to their scaling properties. Thus, the more extreme variations in value-added scores can not be attributed to differences in the measuring capacities of different forms of the norm-referenced part of the TCAP tests.

B. Are there flaws in the software?

No. Great care has been exerted to independently verify that the software produces accurate results. In small test cases, it has been demonstrated that the results from the TVAAS software are identical to results obtained using other software. (Note: commercially available software will only handle the mixed model equations for the smallest school systems, thus small examples are used for verification. The software developed to execute the computation of TVAAS is much more powerful than any commercially available program because it must deal with massive computing requirements far beyond those for which the commercial programs were devised.) Furthermore, validation studies carried out by independent experts in statistics are currently underway. The results of these studies will be available within the next several weeks. However, the first preliminary test results, directed by a statistical expert from another university, confirmed the accuracy of the TVAAS software.

C. Could the deployment of a new educational program affect the value-added scores?

Yes. For example, preliminary results from some schools which incorporated teaching science from a hands-on approach showed cumulative gains consistently above 100%.

D. Could a failure to provide instruction consistent with students' prior level of achievement result in inconsistent value-added scores?

Yes. Even though it varies greatly among systems, the statistical grade*year interaction is highly significantly different from zero for most school systems. This mean that different groups of students enter the classroom with different levels of initial attainment in different years. The greater this interaction, the greater the variation from year to year in the value-added scores.

The state-wide data strongly suggest that if instruction commensurate with the prior level of educational achievement is provided for all students, then the magnitude of this interaction will be greatly reduced. The net result of teaching to

students at the level at which they enter the classroom is that students of all achievement levels will make more appropriate gains and that there will be less variation from year to year.

It has been said, "We did not change a thing, yet our value-added scores changed drastically." This or a similar statement has often been provided to raise questions about the validity of the TVAAS models. The statement can be completely true and yet expose exactly why the value-added scores are observably different among years. Consider the hypothetical case of a school with a static delivery system, **as depicted in figure 1**. The same material, represented by the gray area labeled "Functional Curriculum," is taught to each grade year after year.

Assume that in the first year, when the students in cohort 1 enter this school*grade*subject, these practices exactly match their needs. Then the gains for these students across all achievement levels will be good. Now assume that the same practices are used with cohort 2, which contains more students performing at a lower level. The gains for the lowest achieving students will not be good, resulting in lower overall value-added scores. In cohort 3, which contains a high concentration of high achievers, the gains for these highest achieving students will not be good, also resulting in lower overall value-added scores. This is a conceptual example, but many educators have reported to us that, by inspecting gain by achievement groups, they have been able to identify which end of the distribution of students needs extra attention. When such attention is properly given, variation from year to year will drop.

Three years of TVAAS reports at the system level and two years at the school level have been prepared and released. Educators across the state have begun to react to these reports. If a school's gain in a specific grade and subject was less than desirable as reported by TVAAS, often attention was given, and the effects of the additional emphasis in these areas were reflected in higher gain scores in the subsequent year. Therefore, children leaving a grade where practices had been improved would enter the subsequent grade at a higher achievement level than the previous year's students. Unless compensatory changes were made in the receiving grade, i.e., unless these students were taught from the higher level at which they entered the classroom, their gains would not be as high as gains of students from the previous year because even though they entered the year performing at a higher level, they would receive the same instruction that was tailored for students performing at the previous year's [lower] level.

To sustain consistent value-added scores over time, tremendous planning and coordination across grade levels is a necessity. In the early pilot studies, the data were analyzed retrospectively, and there was either no feed-back to the schools or it was ignored. In either case, the process was rather static. Today, the data are being used to inform educational practice, creating more "waves" in the process. The variation from year to year offers strong evidence that TVAAS is having the desired effects.

Over time, as educators at all levels learn how to more efficiently sustain academic growth for students of all levels, the process will become more stable. However,

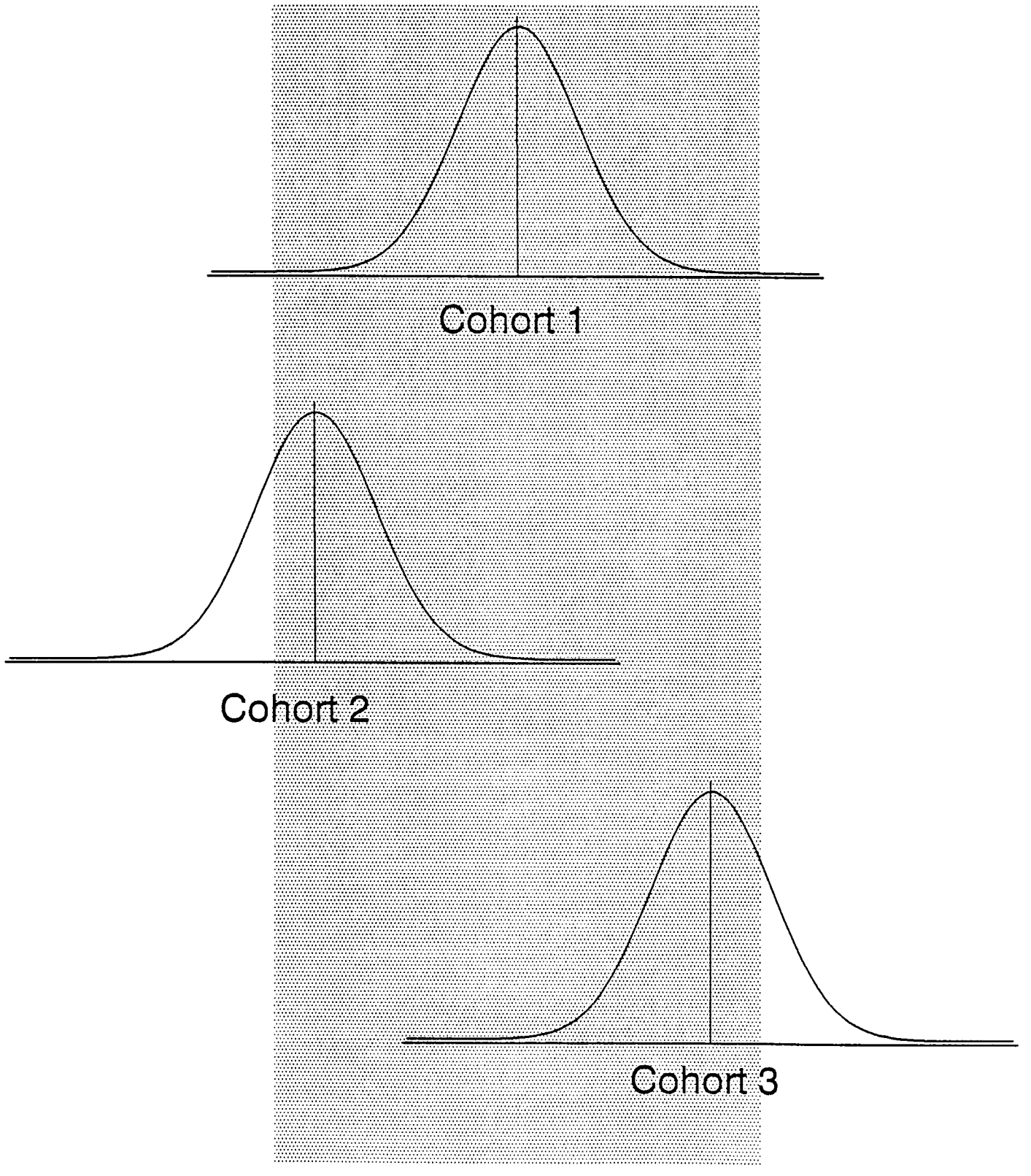

Figure 1.

there will always be some fluctuation from year to year, even under the best circumstances. This is why the 3-year average should receive the most attention. Nevertheless, the magnitude of difference in gains over years reported for the extreme cases cited in the draft of the Office of Education Accountability report can be greatly reduced as evidenced by the relative stability in the gains of hundreds of Tennessee schools which could be cited.

Office of Education Accountability Comment

- The Office of Education Accountability has never claimed that value-added scores should not change. The source of concern has been the large, unexplained, and seemingly random changes.
- Office of Education Accountability analysts have not chosen "extreme cases" to illustrate this point; many other examples could have been cited. Many schools may indeed have stable value-added results—but the existence of a significant number of schools with unstable results raises legitimate questions about the conclusions one can draw from value-added results.
- Sanders believes that variations in gains are a result of either: the failure of schools and systems to adapt their teaching methods to children's instructional needs; or the positive responses of teachers and schools to the information supplied by the TVAAS. Evidence to support these interpretations needs to be gathered.

Factors Affecting Student Gain Have Not Been Identified

(Executive Summary, p. ii; Report Text, pp. 14-17)

This is the most serious error contained in the report. For more than a century, educators have been attempting to agree on the five, twenty-five, or 125 factors that influence student learning. The Report overstates when it implies that "education research" has identified the factors that, taken together, can predict student performance. Indeed, the premise of the value-added assessment lies in the belief, supported by the empirical data, that it is possible to measure student learning in a way that mitigates greatly the influence of demographic factors.

The factors cited may be real, but one does not need a direct measure of them to provide unbiased estimates of the influences of system, schools and teachers on the rate of academic growth. To provide a filter of these exogenous influences, TVAAS uses the child's academic performance over time as a "blocking factor." (Blocking is an old, well-known, and often-used statistical method that enables the measurement of the influence of one or more factors on some response variable, independently of the direct influence of other factors.)

To be more precise, the academic growth of a child over time contains within it the relatively stable factors of socioeconomic status, educational attainment of parents, family structure, community, ability, gender, race, and many other influences. Deviation from the normal academic growth pattern can be influenced by changes in these factors, but our analyses and findings have shown that the major influences on academic growth of children, as determined utilizing mixed-model methodology, are those of schools, school systems, and teachers. For instance, TVAAS has shown student gains to be unrelated to

the percentage of minority students; the socioeconomic status, as determined through the percentage of free and reduced price lunches; or the location of the school—urban, rural, or suburban. Moreover, since educational effects are based on the gains of cohorts of students over a period of three years, the fluctuations one would expect to occur in the lives of students tend to be mitigated over time. Where one student suffers through a divorce, another is removed from an abusive situation.

If the exogenous factors mentioned had a major influence, it would be unreasonable to expect to find many intercity schools with high cumulative value-added scores. In fact many examples of excellent value-added scores can be cited for intercity schools, as well as rural and suburban schools, offering the empirical evidence that TVAAS is indeed providing unbiased estimates as designed.

The blocking aspects (in other words, the multivariate repeated measures aspect of the TVAAS models) have eliminated the need for the "market basket" of covariables that have been advocated by some educational researchers. If the covariables were available, then similar results would be expected. However, a requirement that this vector of information be available for each child insures that such a strategy could never be deployed because of the virtual impossibility of ever having the required complete data record for each child.

Nevertheless, TVAAS results are constantly monitored to detect unexpected bias occasioned by exogenous variables, should it occur. If such influence should be detected, TVAAS allows for the inclusion of such variables without major changes in the model.

The single most striking result of the data collected thus far from the value-added assessment makes a valuable contribution to this discussion. In Tennessee, success or lack of success in student academic gains appear to bear no relation to a school's average income or racial composition. Likewise, average gains do not reflect whether a school is urban, rural or suburban. Each category has numerous examples of schools that exceed and fall short of national academic gains.

There are many Memphis schools with more than 80 percent minority enrollment that exceed national norm gains. They stand in contrast to a host of well-funded suburban schools that fall far short of national norm gains in more than one grade and subject. There are dozens of rural schools with largely white enrollments that exceed national gains, with a comparable number that fall below expectations.

Education researchers who have staked their careers on proving the influence of demographic factors on student learning can be expected to resist the conclusions of the data present in Tennessee. One should note, however, that the national debate over the assumptions of the value-added analysis lists as many supporters as detractors.

Office of Education Accountability Comment

The point made by the Office of Education Accountability report is that the "blocking effect" needs to be evaluated by other experts in the fields of statistics, educational measurement, and testing. This is not to dispute Sanders' claims in this area; it is merely to point out that such claims contradict a great deal of widely-accepted educational measurement theory and require rigorous independent testing.

Scotts Hill Question

(Executive Summary, p. ii-iii; Report Text, pp. 17-18)

The Report's interpretation of the data for this school is simply wrong. The difference in scores between residents of Decatur and Henderson counties is not a consequence of the value-added assessment; rather, the data reflect that Scotts Hill School is more effective in improving the scores of its lower achieving language arts students than its higher achieving students.

A close look at the data shows that a larger percentage of the higher achieving students are from Henderson County. If the curriculum is geared to the lower achieving students, it follows that the larger gains will come from the lower achieving students in Decatur County.

The Scotts Hill example, as presented, is misleading, provides misinformation and misconstrues TVAAS. The conclusions drawn from this example are wrong and provably so. First, our previous research has indicated that no relationship has been found with the AVERAGE entering point of groups of students and their subsequent MEAN gains. This should never be interpreted to imply that within all classrooms or schools that high achievers and low achievers are given the same opportunity to make satisfactory growth. In fact, the differences in opportunity to express academic growth can be rather dramatic within the same group of students.

For the past two years in public presentations, a concentrated effort has been made to display and explain these differences. The terms "tee pee" pattern, "shed" pattern, and "reverse shed" are terms which have been coined to describe the most prevalent pattern of unequal growth among students of differing achievement levels. Last year the gains by achievement groups were calculated and mailed to systems and schools within the state. This year, this "Additional Diagnostic Information" for produced for ALL schools in Tennessee and was forwarded to the Department of Education for distribution.

How does this relate to the case of Scotts Hill School? Clearly, if practices there are most appropriate for students at a particular achievement level, and if students at that level are predominantly from one of the two school systems rather than the other, that system's students will have better gains.

Is there evidence of this? Consider the case of 5th grade gains in Language. The reported TVAAS 3-year-average gains are 39.3 for Decatur County and 26.5 for Henderson County. (The raw mean gains are 35.5 and 27.3, respectively, based on 34 and 50 students in 1992-1994. The difference is NOT a consequence of TVAAS; it is simple a characteristic of the data.) Combining the two counties and breaking down the results by achievement level reveals the following pattern:

Level:	650-699	700-749	750-799
Avg Gain:	36.0	30.8	24.2
N:	19	42	16

Scotts Hill seems to be doing a better job with the students in the lower achievement levels. It turns out that a higher percentage of Decatur County students is in the 650-699 category (32% versus 16% for Henderson County). Roughly the same percentage in each county is in the 700-749 category (47% and 52%). A lower percentage of Decatur County

students are in the 750-799 category (12% versus 24%). The Decatur County students thus have an advantage over the Henderson County students in that more of them are at the achievement level that the educational practices are addressing. Consequently, they get higher gains.

To reiterate, the difference is NOT a consequence of the TVAAS model; it is simply a characteristic of the data which reflects that this school is more effective with its lower achieving language arts students than its higher achieving ones in this grade. Findings such as these are the very reason that this additional information has been provided to the Department for distribution. Many, many educators have reported that they have found that this information has been most helpful in targeting those areas that need special attention.

The Scotts Hill data tend to show that TVAAS is doing what it was designed to do. The official TVAAS report reflects what the group average has gained, not specific subgroups. Thus, the alleged discrepancy between the "two" Scotts Hill school reports is not a discrepancy but rather reflects the differential success of Scotts Hill to invoke gains at a different rate over its distribution of students.

Office of Education Accountability Comment

Office of Education Accountability analysts have never claimed that the differences between the two groups of students are a consequence of the model. The interpretation of those differences is at issue.

Using the data supplied by Sanders, Office of Education Accountability analysts calculated projected gain scores by student distributions (i.e., high, middle, and low-achieving student groups).

Assumption 1

This calculation is based on the assumption that the lowest achieving group got the highest average gain and the highest achieving group got the lowest average gain.

Scotts Hill 5th Grade Language Scores

		Both Counties			Decatur County Students			Henderson County Students		
Achievement Groups	**Avg Gain 1992-94**	**Total N**	**%**	**Wtd Gain**	**n**	**%**	**Wtd Gain**	**n**	**%**	**Wtd Gain**
under 650	***42***	4	5%	2	2	6%	2.47	2	4%	1.68
650-699	36.0	19	23%	8.14	11	32%	11.65	8	16%	5.76
700-749	30.8	42	50%	15.40	16	47%	14.49	26	52%	16.02
750-799	24.2	16	19%	4.61	4	12%	2.85	12	24%	5.81
800 or more	***18***	3	4%	0.64	1	3%	0.53	2	4%	0.72
Total		84	100%	**30.80**	34	100%	**31.99**	50	100%	**29.98**

Assumption 2

This calculation is based on the data that Sanders provided in his response. Average gain for the group scoring under 650 is assumed to be equal to the average gain for the group scoring between 650 and 699. Average gain for the

group scoring over 800 is assumed to be equal to the average gain for the group scoring between 750 and 799.

Scotts Hill 5th Grade Language Scores

		Both Counties			Decatur County Students			Henderson County Students		
Achievement Groups	**Avg Gain 1992-94**	**Total N**	**%**	**Wtd Gain**	**n**	**%**	**Wtd Gain**	**n**	**%**	**Wtd Gain**
under 650	***36***	4	5%	1.71	2	6%	2.12	2	4%	1.44
650-699	36.0	19	23%	8.14	11	32%	11.65	8	16%	5.76
700-749	30.8	42	50%	15.40	16	47%	14.49	26	52%	16.02
750-799	24.2	16	19%	4.61	4	12%	2.85	12	24%	5.81
800 or more	***24.2***	3	4%	0.86	1	3%	0.71	2	4%	0.97
Total		84	100%	**30.73**	34	100%	**31.82**	50	100%	**29.99**

Assumption 3

This calculation is based on the three groups that Sanders provided with average gains. Those three groups together accounted for 92 percent of actual students—77 out of 84.

Scotts Hill 5th Grade Language Scores

		Both Counties			Decatur County Students			Henderson County Students		
Achievement Groups	**Avg Gain 1992-94**	**Total N**	**%**	**Wtd Gain**	**n**	**%**	**Wtd Gain**	**n**	**%**	**Wtd Gain**
650-699	36.0	19	24.68%	8.88	11	35%	12.77	8	17%	6.26
700-749	30.8	42	54.55%	16.80	16	52%	15.90	26	57%	17.41
750-799	24.2	16	20.78%	5.03	4	13%	3.12	12	26%	6.31
Total		77	100%	**30.71**	31	100%	**31.79**	46	100%	**29.98**

Summary

Scotts Hill 5th Grade Language Scores--Summary

		Both Counties		Decatur County		Henderson County	
Achievement Groups	**Avg Gain 1992-94**	**Total N**	**%**	**n**	**%**	**n**	**%**
<650		4	5%	2	6%	2	4%
650-699	36.0	19	23%	11	32%	8	16%
700-749	30.8	42	50%	16	47%	26	52%
750-799	24.2	16	19%	4	12%	12	24%
>799		3	4%	1	3%	2	4%
Total		84	100%	34	100%	50	100%

	Both Counties	Decatur County	Henderson County	Difference
Actual TVAAS Results		39.3	26.5	12.8
Raw Mean Gains		35.5	27.3	8.2
Projected Gains w/ 1st assumption	30.8	32.0	30.0	2.0
Projected Gains w/ 2nd assumption	30.7	31.8	30.0	1.8
Projected Gains w/ 3rd assumption	30.7	31.8	30.0	1.8

Sanders states: “The Decatur County students thus have an advantage over the Henderson County students in that more of them are at the achievement level that the educational practices are addressing. Consequently, they get higher gains." Office of Education Accountability analysts disagree; there is no evidence to support this hypothesis.

It is true that a higher percentage of students from Decatur were in a lower-achieving subgroup than students from Henderson. However, if students within subgroups had similar gains—as Sanders believes they had—this distribution difference would translate into a very small difference in gains.

The value-added difference of 12.8 and the raw mean gain diffence of 8.2 were far larger than the projected difference of two (2) points or less caused by achievement group differences. Sanders believes that the Office of Education Accountability analysis is not valid because of the small number of cases; it deserves further study by an independent third party.

Limited Number of People Who Know About Value-Added Software

(Executive Summary, p. iii; Report Text, pp. 18-19)

Written procedures for the value-added assessment software exist. These guidelines were designed specifically for use by those who run the system, now or in the future. The operation of the value-added assessment program is not contingent upon the presence of any single person or group of persons.

Written procedures are in place for those components of the software system which are complete. Dr. Saxton has carefully documented the procedure for processing the mixed-model part of the system. Mr. Schneider has written documents for the data editing-merging step. Mr. Wright has documentation that details the report generating component of the process. Additionally, members of the TVAAS software development team are familiar with and have run the software developed by other team members.

Office of Education Accountability Comment

Written procedures had not been developed at the time Office of Education Accountability analysts visited VARAC. The Office of Education Accountability has requested that the Division of State Audit perform an Information Systems Assessment. This should alleviate any concerns in this area.

TVAAS Is Not Easily Explainable

(Executive Summary, p. iii; Report Text, p. 19)

Explaining the TVAAS in a simple way continues to be a challenge. More than 50 meetings have been conducted with teachers and administrators to explain the value-added assessment and how is results can be used to identify strengths and weaknesses in a particular school or system. These meetings have been attended by thousands of Tennessee teachers, principals, supervisors, superintendents, and school board members who have received answers to a host of legitimate questions.

A majority of teachers and administrators have sought to learn more about the model and make suggestions to improve its implementation. As expected the Report had no difficulty finding voices opposed to the value-added assessment.

These voices confirm a prediction made by those who developed the model and by those in the Legislature who voted for its inclusion in the 1992 reforms. Among schools and systems that do not score well on portions of the assessment, there are two distinct reactions among teachers and administrators. In the first group, a sincere desire exists to identify areas of weakness, determine their cause, and develop a plan to correct them. Success stories abound of systems that have addressed their problems in a forthright manner. Those who administer the model are obligated to assist these educators in every way possible.

In contrast, the second group spends much of its energy in denial, refusing to acknowledge the possibility that its students might be capable of performing better in a particular subject or grade. Not surprisingly, they make little effort either to understand the TVAAS process or benefit from the information it provides. To this group, no apology is warranted.

Much effort has been expended by this office to inform educators and the general public about the TVAAS model. In addition to workshops, meetings with various constituency groups, scores of presentations, dozens of video tapes, and written and verbal communications with administrators, teachers, legislators, business leaders, and many others, TVAAS has a monthly column in the *TEA News* where teacher questions are answered. The first publication on TVAAS was a brochure designed to answer teacher questions. Nevertheless, it is true that many educators still do not understand the principles behind TVAAS.

The level of understanding and utility not only of the TVAAS information but test information in general varies enormously among systems. In those systems in which the local leadership has made assertive attempts to provide detailed explanations to its faculties, then the level of understanding is high and the degree of apprehension is much lower. However, a concerted effort from UT-VARAC, UT-STEC, Department of Education, colleges of education and local school administrators will be necessary to optimize the understanding and utilization of the power of TVAAS.

As with any new, unprecedented process, growing pains are and will be evident. The naysayers, including some members of the Department of Education, have contributed to some of the misgivings. For instance, until the reorganization of State Testing a little over a year ago, the consultants used by that organization to inform teachers about testing and test interpretation were directed to respond to questions on TVAAS by stating that they knew nothing about it.

Recently, this policy has been changed. It is now recognized that the totality of the information that TVAAS and TCAP can provide to educators is remarkable and a comprehensive tool for improving educational practice in Tennessee. TCAP criterion-referenced tests provide detailed information on the mastery of skills by individual students while TVAAS provides more general information on the academic progress of groups of

students. The two together can give educators varied means of assessing the efficacy of educational strategies.

The broader the acceptance of TVAAS then the greater its usefulness will be. This will occur more rapidly as more entities concerned with education recognize the value of the information TVAAS provides and assist in informing educators how to effectively use it. Meanwhile, efforts to reach as many educators as possible for the purpose of increasing their understanding of value-added assessment will continue from UT-VARAC.

High Stakes Nature of TCAP

(Executive Summary, p. iii; Report Text, pp. 19-21)

The real incentives for our students to perform well far outweigh the issues of sanctions or financial rewards to schools and systems. There is no evidence of students deliberately scoring poorly in order to harm a teacher, or of teachers illegally administering the TCAP test. Even the presence of isolated examples would not be sufficient to invalidate the results of a process that provides accountability to those who pay for the cost of public education.

Certainly, the EIA provides strict punishments for anyone found to be compromising the integrity of the tests or the testing procedures...Further, each school system must have on file with the Department of Education a written plan, detailing how test security will be assured. Common means of protecting test security are the use of proctors and the rotating of teachers from their own to another teacher's class for the monitoring of the testing.

As to the "teaching to the test" phenomenon, your example of the principal who was planning to make changes reducing "teaching to the test" already observed within his school on the basis of what he had learned from TVAAS reports is precisely what is hoped will occur. Because of TVAAS, fresh, non-redundant equivalent tests are mandated each year. Since the tests consist primarily of new items, teaching to the test becomes extremely difficult if not impossible.

In regards to classifying low ability or low achieving students as special education to remove their scores from a teacher's TVAAS assessment, there is no efficacy to this practice, should it ever occur. Low achieving students are at least as likely as their higher achieving classmates to make normal gains as evidenced by five years of data collected for TVAAS. Furthermore, the scores of these students *are* included in the assessment of schools and systems Teachers who are responsible for these students will contribute to the overall estimate of school effects, even though the scores are not used for the teacher reports. TVAAS reports for the schools and systems will indicate whether these students are being effectively taught and, if not, it is likely that more attention will be given to their appropriate instruction.

<u>Office of Education Accountability Comment</u>

- "Teaching to the test" can take many forms, including spending instructional time to teach test-taking skills or ignoring areas of the state curriculum that are not reflected in a nationally norm-referenced test (Tennessee history is a good example). The Office of Education Accountability agrees that the statute

requiring “fresh, non-redundant” tests each year would prevent teaching to test *items*; but the concern remains that high stakes tests for teachers may lead to counterproductive use of instructional time. Educators need encouragement and leadership from the State Board of Education and the State Department of Education; these organizations could provide information and support to schools and districts wishing to move away from the types of coaching strategies that are counterproductive.

- Whether or not special education students take the TCAP depends on their Individual Education Plan (IEP). It is possible that special education students might not take the test at all—and thus, their gains would not be reflected in the school or district scores. This issue needs to be monitored carefully.
- The State Department of Education has documented evidence of at least one student who attempted to organize his classmates to do poorly on the TCAP.